Health Technology Assessment 2007; Vol. 11: No. 3

A systematic review of rapid diagnostic tests for the detection of tuberculosis infection

J Dinnes, J Deeks, H Kunst, A Gibson,
E Cummins, N Waugh, F Drobniewski
and A Lalvani

January 2007

Health Technology Assessment
NHS R&D HTA Programme
www.hta.ac.uk

How to obtain copies of this and other HTA Programme reports.
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

A systematic review of rapid diagnostic tests for the detection of tuberculosis infection

J Dinnes,[1*] J Deeks,[2] H Kunst,[3] A Gibson,[4] E Cummins,[5] N Waugh,[6] F Drobniewski[4] and A Lalvani[7]

[1] Wessex Institute for Health Research and Development, University of Southampton, UK
[2] Centre for Statistics in Medicine, University of Oxford, UK
[3] Department of Respiratory Medicine, Royal Brompton Hospital, London, UK
[4] HPA National Mycobacterium Reference Unit, London, UK
[5] McMaster Development Consultants, Glasgow, UK
[6] Department of Public Health, University of Aberdeen, UK
[7] Nuffield Department of Clinical Medicine, University of Oxford, UK

* Corresponding author

Declared competing interests of authors: A Lalvani is a named inventor on patents relating to T cell-based diagnosis filed by the University of Oxford. Regulatory approval and commercialisation of ELISPOT (T-SPOT TB) has been undertaken by a spin-out company of the University of Oxford (Oxford Immunotec Ltd), in which AL has a share of equity and to which he acts as a non-executive director and scientific advisor

Published January 2007

This report should be referenced as follows:

Dinnes J, Deeks J, Kunst H, Gibson A, Cummins E, Waugh N, *et al*. A systematic review of rapid diagnostic tests for the detection of tuberculosis infection. *Health Technol Assess* 2007;**11**(3).

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

NIHR Health Technology Assessment Programme

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

The research findings from the HTA Programme directly influence decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC). HTA findings also help to improve the quality of clinical practice in the NHS indirectly in that they form a key component of the 'National Knowledge Service'.

The HTA Programme is needs-led in that it fills gaps in the evidence needed by the NHS. There are three routes to the start of projects.

First is the commissioned route. Suggestions for research are actively sought from people working in the NHS, the public and consumer groups and professional bodies such as royal colleges and NHS trusts. These suggestions are carefully prioritised by panels of independent experts (including NHS service users). The HTA Programme then commissions the research by competitive tender.

Secondly, the HTA Programme provides grants for clinical trials for researchers who identify research questions. These are assessed for importance to patients and the NHS, and scientific rigour.

Thirdly, through its Technology Assessment Report (TAR) call-off contract, the HTA Programme commissions bespoke reports, principally for NICE, but also for other policy-makers. TARs bring together evidence on the value of specific technologies.

Some HTA research projects, including TARs, may take only months, others need several years. They can cost from as little as £40,000 to over £1 million, and may involve synthesising existing evidence, undertaking a trial, or other research collecting new data to answer a research problem.

The final reports from HTA projects are peer-reviewed by a number of independent expert referees before publication in the widely read monograph series *Health Technology Assessment*.

Criteria for inclusion in the HTA monograph series

Reports are published in the HTA monograph series 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 referees 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.

The research reported in this monograph was commissioned by the HTA Programme as project number 01/02/07. The contractual start date was in May 2002. The draft report began editorial review in January 2006 and was accepted for publication in February 2006. As the funder, by devising a commissioning brief, the HTA Programme specified the research question and study design. 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 referees 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.

The views expressed in this publication are those of the authors and not necessarily those of the HTA Programme or the Department of Health.

ISSN 1366-5278

Published by Gray Publishing, Tunbridge Wells, Kent, on behalf of NCCHTA.
Printed on acid-free paper in the UK by St Edmundsbury Press Ltd, Bury St Edmunds, Suffolk.

Abstract

A systematic review of rapid diagnostic tests for the detection of tuberculosis infection

J Dinnes,[1*] J Deeks,[2] H Kunst,[3] A Gibson,[4] E Cummins,[5] N Waugh,[6] F Drobniewski[4] and A Lalvani[7]

[1] Wessex Institute for Health Research and Development, University of Southampton, UK
[2] Centre for Statistics in Medicine, University of Oxford, UK
[3] Department of Respiratory Medicine, Royal Brompton Hospital, London, UK
[4] HPA National Mycobacterium Reference Unit, London, UK
[5] McMaster Development Consultants, Glasgow, UK
[6] Department of Public Health, University of Aberdeen, UK
[7] Nuffield Department of Clinical Medicine, University of Oxford, UK
* Corresponding author

Objectives: To evaluate the effectiveness of available rapid diagnostic tests to identify tuberculosis (TB) infection.
Data sources: Electronic databases were searched from 1975 to August 2003 for tests for active TB and to March 2004 for tests for latent tuberculosis infection (LTBI).
Review methods: Studies were selected and evaluated that (1) tested for LTBI, (2) compared tuberculin skin test (TST) and interferon-γ assays based on ESAT-6 and CFP-10 antigens and (3) provided information on TB exposure or bacille Calmette–Guérin (BCG) vaccination or HIV status. For each test comparison, the sensitivity, specificity and 95% confidence intervals (CIs) were calculated. Sources of heterogeneity were investigated by adding covariates to the standard regression model. The authors examined whether interferon-γ assays were more strongly associated with high versus low TB exposure than TST. Odds ratios (ORs) were calculated for the association between test results and exposures from each study along with their 95% CIs. Within each study, the OR value for one test was divided by that for another to produce a ratio of OR (ROR).
Results: A total of 212 studies were included, providing 368 data sets. A further 19 studies assessing fully automated liquid culture were included. Overall, nucleic acid amplification test (NAAT) accuracy was far superior when applied to respiratory samples as opposed to other body fluids. The better quality in-house studies, were, for pulmonary TB, much better at ruling out TB than the commercial tests (higher sensitivity), but were less good at ruling it in (lower specificity), but it is not possible to recommend any one over another owing to a lack of direct test comparisons. The specificity of NAAT tests was high when applied to body fluids, for example for TB meningitis and pleural TB, but sensitivity was poor, indicating that these tests cannot be used reliably to rule out TB. High specificity estimates suggest that NAAT tests should be the first-line test for ruling in TB meningitis, but that they need to be combined with the result of other tests in order to rule out disease. Evidence for NAAT tests in other forms of TB and for phage-based tests is significantly less prolific than for those above and further research is needed to establish accuracy. There is no evidence to support the use of adenosine deaminase (ADA) tests for diagnosis of pulmonary TB; however, there is considerable evidence to support their use for diagnosis of pleural TB and to a slightly lesser extent for TB meningitis. Anti-TB antibody test performance was universally poor, regardless of type of TB. Fully automated liquid culture methods were superior to culture on solid media, in terms of their speed and their precision. In total, 13 studies were included. Assays based on RD1 specific antigens, ESAT-6 or CFP-10, correlate better with intensity of exposure, and therefore are more likely than TST/purified protein derivative (PPD)-based assays to detect LTBI accurately. An additional advantage is that they are more likely to be independent of BCG vaccination status and HIV status.
Conclusions: The NAAT tests provide a reliable way of increasing the specificity of diagnosis (ruling in

disease) but sensitivity is too poor to rule out disease, especially in smear-negative (paucibacillary) disease where clinical diagnosis is equivocal and where the clinical need is greatest. For extra-pulmonary TB, clinical judgement has both poor sensitivity and specificity. For pleural TB and TB meningitis, adenosine deaminase tests have high sensitivity but limited specificity. NAATs have high specificity and could be used alongside ADA (or interferon-γ) to increase sensitivity for ruling out disease and NAAT for high specificity to rule it in. All studies from low-prevalence countries strongly suggest that the RD1 antigen-based assays are more accurate than TST- and PPD-based assays for diagnosis of LTBI. If their superior diagnostic capability is found to hold up in routine clinical practice, they could confer several advantages on TB control programmes. Further research for active TB needs to establish diagnostic accuracy in a wide spectrum of patients, against an appropriate reference test, and avoiding the major sources of bias. For LTBI, research needs to address different epidemiological and clinical settings, to evaluate the performance of the main existing commercial assays in head-to-head comparison in both developed and developing countries, and to assess the role of adding more TB-specific antigens to try to improve diagnostic sensitivity.

Contents

List of abbreviations

ADA	adenosine deaminase
AFB	acid-fast bacilli
AMTD	Amplified Mycobacterium Direct Test
ATB	active tuberculosis
BAL	bronchoalveolar lavage
BCG	bacille Calmette–Guérin
CFP-10	culture filtrate protein 10
CI	confidence interval
CSF	cerebrospinal fluid
CT	computed tomography
DARE	Database of Abstracts of Reviews of Effects
DOR	diagnostic odds ratio
ELISA	enzyme-linked immunosorbent assay
ELISPOT	enzyme-linked immunospot assay
ESAT-6	early secretory antigen target 6
FALC	fully automated liquid culture
FNA	fine-needle aspirate
FNAC	fine-needle aspiration cytology
FPR	false positive rate
HIV	human immunodeficiency virus
HPLC	high-performance liquid chromatography
HRCT	high-resolution computed tomography
ICER	incremental cost-effectiveness ratio
INH	isoniazid
IVU	intravenous urography
LAM	lipoarabinomannan
LTB	latent tuberculosis
LJ	Lowenstein–Jensen
LTBI	latent tuberculosis infection
MDR-TB	multidrug-resistant tuberculosis
MGIT	mycobacterial growth indicator tube
MRI	magnetic resonance imaging
M. AC	*Mycobacterium avian complex*
M. TB	*Mycobacterium tuberculosis*
NAAT	nucleic acid amplification test
NBM	non-tuberculous mycobacteria
NICE	National Institute for Health and Clinical Excellence
NPV	negative predictive valve
NTM	non-tuberculous mycobacteria
PBMC	peripheral blood mononuclear cells
PCR	polymerase chain reaction
PPD	purified protein derivative

continued

List of abbreviations *continued*

PPV	positive predictive valve
PRA	PCR restriction enzyme analysis
OR	odds ratio
QALY	quality-adjusted life-year
QoL	quality of life
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
RDOR	relative diagnostic odds ratio
RMP	rifampin
RNA	ribonucleic acid
ROC	receiver operating characteristic
ROR	ratio of odds ratios
SD	standard deviation
SROC	summary receiver operating characteristic
TB	tuberculosis
TBLN	tuberculous lymphadenitis
TBSA	tuberculostearic acid
TNF	tumour necrosis factor
TST	tuberculin skin test
TU	tuberculin unit
WHO	World Health Organization

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices in which case the abbreviation is defined in the figure legend or at the end of the table.

Executive summary

Background

Globally, there are 8 million new tuberculosis (TB) cases and 2 million deaths per year. Once infected, active disease develops in about 10% of cases, usually within 1–2 years after exposure. Remaining individuals enter into a state of latentcy [latent tuberculosis infection (LTBI)], which can reactivate at a later stage, particularly if the individual becomes immunocompromised.

Active TB is predominantly pulmonary in nature. Extra-pulmonary TB occurs in approximately 41% of TB cases in England and Wales and includes lymphatic, pleural, meningeal, pericardial, skeletal, gastrointestinal, genitourinary and miliary TB. LTBI has no clinical manifestations and is not contagious.

Given the infectious nature of pulmonary TB, fast and accurate diagnosis is an important element of TB treatment and control.

Objectives

1. For each form of active tuberculosis, to conduct systematic reviews to evaluate the accuracy of the following groups of tests in patients suspected of active TB:
   (a) nucleic acid amplification tests
   (b) amplification molecular probe tests
   (c) serodiagnostic and biochemical assays
   (d) phage-based tests.
2. To conduct a systematic review to evaluate how effective fully automated liquid culture systems are for isolating and identifying TB.
3. To conduct a systematic review to evaluate the use of interferon-γ assays for detection of latent TB infection.
4. To examine the likely NHS and societal consequences of false-positive and-false negative tests.

Methods

Data sources

Literature was identified from electronic databases and other sources. All databases were searched from 1975 to August 2003 for tests for active TB and to March 2004 for tests for LTBI. Reference lists of included studies and relevant review articles were scanned for additional studies.

Study selection

Tests for active TB

Any study comparing a **rapid** test for detection of active tuberculosis with any reference standard was included. 'Rapid' tests were those for which a result could be obtained in less than the time taken for standard culture. Only case series studies were included. Accuracy studies had to report sufficient information to allow the construction of a 2×2 contingency table.

Tests for latent TB infection

The study selection criteria were (1) testing for LTBI, (2) comparison between tuberculin skin test (TST) and interferon-γ assays based on ESAT-6 and CFP-10 antigens and (3) information on TB exposure or bacille Calmette–Guérin (BCG) vaccination or HIV status.

Data extraction

Data extraction and study quality assessment were undertaken independently by two reviewers.

Data synthesis

Tests for active TB

For each test comparison, the sensitivity, specificity and 95% confidence intervals (CIs) were calculated. The method proposed by Moses and colleagues to fit both symmetric and asymmetric summary receiver operating characteristic (SROC) curves was used. Sources of heterogeneity were investigated by adding covariates to the standard regression model.

Tests for latent TB infection

Interferon-γ assays were examined to establish whether they were more strongly associated with high versus low TB exposure than TST. Odds ratios (ORs) were calculated for the association between test results and exposures from each study along with their 95% CIs. Within each study, the OR value for one test was divided by that for another to produce a ratio of OR (ROR).

Results

Tests for active TB

A total of 212 studies were included, providing 368 data sets. A further 19 studies assessing fully automated liquid culture were included.

Overall, nucleic acid amplification test (NAAT) accuracy was far superior when applied to respiratory samples as opposed to other body fluids. The better quality in-house studies were, for pulmonary TB, much better at ruling out TB than the commercial tests (higher sensitivity), but were less good at ruling it in (lower specificity), but it is not possible to recommend any one over another owing to a lack of direct test comparisons.

The specificity of NAAT tests was high when applied to body fluids, for example for TB meningitis and pleural TB, but sensitivity was poor, indicating that these tests cannot be used reliably to rule out TB. High specificity estimates suggest that NAAT tests should be the first-line test for ruling in TB meningitis, but that they need to be combined with the result of other tests in order to rule out disease. Evidence for NAAT tests in other forms of TB and for phage-based tests is significantly less prolific than for those above and further research is needed to establish accuracy.

There is no evidence to support the use of adenosine deaminase (ADA) tests for diagnosis of pulmonary TB; however, there is considerable evidence to support their use for diagnosis of pleural TB and to a slightly lesser extent for TB meningitis.

Anti-TB antibody test performance was universally poor, regardless of type of TB. Fully automated liquid culture methods were superior to culture on solid media, in terms of their speed and their precision.

Tests for latent TB infection

In total, 13 studies were included. Assays based on RD1-specific antigens, ESAT-6 or CFP-10, correlate better with intensity of exposure, and therefore are more likely than TST/purified protein derivative (PPD)-based assays to detect LTBI accurately. An additional advantage is that they are more likely to be independent of BCG vaccination status and HIV status.

Conclusions

Implications for healthcare

The NAAT tests provide a reliable way of increasing the specificity of diagnosis (ruling in disease) but sensitivity is too poor to rule out disease, especially in smear-negative (paucibacillary) disease where clinical diagnosis is equivocal and where the clinical need is greatest.

For extra-pulmonary TB, clinical judgement has both poor sensitivity and specificity. For pleural TB and TB meningitis, adenosine deaminase tests have high sensitivity but limited specificity. NAATs have high specificity and could be used alongside ADA (or interferon-γ) to increase sensitivity for ruling out disease and NAAT for high specificity to rule it in.

All studies from low-prevalence countries strongly suggest that the RD1 antigen-based assays are more accurate than TST- and PPD-based assays for diagnosis of LTBI. If their superior diagnostic capability is found to hold up in routine clinical practice, they could confer several advantages on TB control programmes.

Recommendations for research

Active TB

Diagnostic accuracy must be established, preferably prospectively, in a wide spectrum of patients, against an appropriate reference test, and avoiding the major sources of bias such as verification bias, lack of blinding, and inclusion of all indeterminate results.

- For pulmonary TB, a study of the accuracy of NAAT in clinically equivocal smear-negative patients is needed, to identify how high a proportion of false-positive results would be generated in this population.
- The place of ADA, interferon-γ and lysozyme for diagnosis of pleural TB requires further investigation
- The place of ADA, for diagnosis of TB meningitis, needs to be established
- For both pleural and TBM, the combination of NAAT tests with other tests such as ADA should be examined
- The incremental value of combinations of tests, particularly for samples of biological fluids, needs assessment in large, prospective, well-designed studies recruiting representative samples of patients.

Interferon-γ assays for the rapid identification of latent tuberculosis infection

- Research is needed in different epidemiological and clinical settings, not only in developed countries, but also in developing countries, and countries with a high prevalence of TB, of non-tuberculous mycobacteria, in populations with

high BCG coverage and in immunosuppressed populations.
- Trials to evaluate the performance of the main existing commercial assays [whole blood interferon-γ enzyme-linked immunosorbent assay (ELISA) and enzyme-linked immunospot assay (ELISPOT)] in head-to-head comparison should be done in both developed and developing countries.
- The role of adding more TB-specific antigens to try to improve diagnostic sensitivity needs to be assessed.
- Longitudinal cohort studies to confirm the positive predictive value of interferon-γ assays for subsequent development of active TB should also be performed.

Chapter 1
Background

Introduction

Tuberculosis (TB) is a major cause of morbidity and mortality throughout the world. One-third of the world's population is infected with the TB bacillus.[1] Mycobacteria are aerobic, non-spore forming, non-motile, single-cell bacteria. Of more than 40 currently recognised species of mycobacteria, *Mycobacterium tuberculosis* (*M. TB*) is the leading cause of death world-wide that can be attributed to a single infectious disease agent. TB is the disease caused by the bacteria of the *M. TB* complex: *M. TB*, *M. bovis* and *M. africanum*. Other mycobacteria can also cause disease or diagnostic problems, including the atypical forms such as *M. avium* species. *M. avium complex* (*M. AC*) disease occurs either as a disseminated disease largely in patients with HIV infection, or as a pulmonary disease in immunocompromised patients. The rapidly growing atypical mycobacteria, including *M. fortuitum*, *M. chelonae* and *M. abscessus*, cause cutaneous, pulmonary and postsurgical wound infections.[2]

The WHO cites TB as the single most important fatal infection, with around 8 million new cases and 2 million deaths per year, 95% in developing countries.[3] Once infected, active disease develops in about 10% of cases, usually within 1–2 years after exposure from TB.[4] The remainder stay in a state of latent tuberculosis infection (LTBI), which can reactivate at a later stage, particularly if the individual is elderly or becomes immunocompromised. There is currently no certain way of predicting which 10% of infected persons will go on to present with active disease, and this remains one of the major challenges in TB research.

When mycobacterial disease is suspected, the aim of testing is to identify the presence of mycobacteria, to characterise isolates and to determine their antibiotic susceptibilities.

Tuberculosis in the UK

TB has never been eradicated in Western societies, but it has been suppressed by public health measures such as selective screening, surveillance and follow-up. Improved living conditions, including nutrition, and introduction of chemotherapy have led to steadily declining disease rates at least in developed countries. Since 1985, TB has, however, re-emerged as a major public health concern with disease rates climbing world-wide.[5,6] In many developed countries such as North America, disease rates continue to decline, whereas in the UK rates are on the incline.[7] Increasing incidence rates have led to the publication of the both the TB Action plan[8] and National Institute for Health and Clinical Excellence (NICE) guidelines on TB.[9]

The recent outbreaks in North London, Leicester and Wales highlight that TB is still a problem in the UK.[10–12] Currently, in the UK there are approximately 6700 cases of clinical TB per annum, with varying geography. This represents a rate of 12.4 per 100,000 population (England, 13.0; Wales, 6.3; Northern Ireland, 3.3).[13] About 40% come from London, which accounts for over two-thirds of the annual increase in notifications in the UK.[13] TB in London is largely caused by reactivation or importation of infection by recent immigrants.[14] In 2003, the mortality rate from TB in England and Wales was around 0.74 per 100,000 population, varying from under 2% for children and young adults to over 30% in those aged 65 years or over.[13]

The rates of new infections have been increased by the marked rise in international travel, by the breakdown of public health measures in some of the Eastern European countries and by increases in the number of people with impaired immunity following the spread of HIV infection.[2] In most developed countries, TB mostly affects older people, recent immigrants from developing countries, members of ethnic minorities and the immunocompromised (mainly HIV). Other populations at risk for developing TB are those with diabetes mellitus,[15] those on immunosuppressive medication post organ transplantation[16] and populations receiving treatment with tumour necrosis factor alpha (TNF-α) antagonists for rheumatoid arthritis and other autoimmune diseases.[17]

HIV infection substantially increases the risk of developing TB once infected with the bacillus, and also shortens the time to development of the disease.[18] Those with double infection have an estimated 10% risk of developing active TB each year.[3,19] HIV-positive patients may be at 10 times greater risk of multidrug-resistant tuberculosis (MDR-TB) than HIV-negative patients.[20] Tuberculosis in HIV-infected individuals may have unusual features, such as atypical pulmonary manifestations or false-negative microbiological results, which can cause diagnostic difficulties.[18]

Clinical manifestations of tuberculosis

Clinical, laboratory and radiological diagnosis of tuberculosis

Delays of up to 2–3 weeks in the management of patients with active TB, including delays in initial suspicion and in initiation of antituberculous chemotherapy, are common.[21] Longer delays frequently occur and delayed diagnosis is usually an important contributory factor in fatal cases of TB in the UK. Physical examination is of limited value in diagnosing pulmonary TB as signs are non-specific. Often results of pulmonary examinations are normal. Physical examination is, however, important in extrapulmonary forms of tuberculosis, namely cardiovascular examination to detect signs of pericarditis and neurological examination to detect meningitis or cord compression due to spinal TB.[22] Laboratory signs such as elevated sedimendation rate, C-reactive protein, leucocytosis, lymphopenia and anaemia may be helpful, but are not diagnostic of TB.[23] Chest radiography is often used in the diagnosis of patients with active TB, as infection usually leads to characteristic radiological patterns including cavities, hilar and mediastinal lymphadenopathy and upper lobe infiltrates.[23] Since a chest radiograph may also show signs of past infection, TB cannot be diagnosed with certainty from the chest radiograph alone, but needs microbiological confirmation. High-resolution computed tomography (HRCT) thorax is reported to be more sensitive than chest radiography and has shown to be useful in predicting the activity of TB, but, like other radiological tests for TB, generally lacks specificity.[24] Hence there is a great and unmet need for sensitive, specific and rapid diagnostic tests for TB.

Pulmonary tuberculosis

Active TB predominantly takes the form of pulmonary TB and can be either primary or postprimary. Primary pulmonary TB occurs soon after initial infection, whereas postprimary TB occurs in the previously infected person as a result of endogenous reactivation of a latent infection or of exogenous reinfection.[6,25] In countries where TB is prevalent, the primary form caused by infection of a non-immune host usually occurs in childhood. In countries where TB is less common, this may occur in adulthood. Once active, the course of pulmonary TB is highly variable and symptoms can vary from an insidious mild, persistent cough to an acute pneumonia-type syndrome, to an ongoing debilitating process with chronic cough, haemoptysis, fevers, night sweats and weight loss.[26] However, some patients with active TB may even be asymptomatic.[27] LTBI has no clinical manifestations and it is neither contagious nor harmful.

Extrapulmonary tuberculosis

In England and Wales, 41% of TB cases were reported to have TB involving extrapulmonary sites only.[13] Extrapulmonary TB was most frequent in the 25–44 years age groups. This may be explained by the fact that foreign-born persons are twice as likely to have extrapulmonary TB than those born in the UK and they represent a large proportion of TB cases in these age groups.[13] It has been estimated that in England and Wales, 23% of white patients and 43% of patients of Indian subcontinent origin present with extrapulmonary TB.[28] The commonest extrapulmonary manifestation of TB is lymphatic TB. Other forms of extrapulmonary TB include pleural, meningeal, pericardial, skeletal, gastrointestinal, genitourinary and miliary TB.[23] The most frequent extrapulmonary sites of disease seen in England and Wales were extrathoracic lymph nodes (16%), intrathoracic lymph nodes (7%), pleura (7%) and joint and bones (5%). Meningeal tuberculosis was reported in 1.5%.[13]

Tuberculous lymphadenitis

Tuberculous lymphadenitis (TBLN) occurs most commonly in cervical lymph nodes (75–90%).[29–31] In resource-poor countries, TBLN is usually a clinical diagnosis supported by a positive tuberculin skin test (TST). The laboratory diagnosis depends on the detection of *M. TB* in pus, aspirates or biopsies from lymph nodes. In children, isolation of *M. TB* is important since infection with non-tuberculous mycobacteria can also give rise to lymphadenopathy. Fine needle aspirate (FNA) cytology showing characteristic morphological findings[24] and smear examination for acid-fast bacilli (AFB) has assumed an important role in the diagnosis of peripheral

lymphadenopathy. It is less invasive than excision biopsy and fine needle aspiration has shown yields of AFB detection of up to 60%.[32,33] Expertise in excisional biopsy is often limited in developing countries.[34] Histology of excisional biopsy shows in over 90% evidence of caseous granulomas.[35] Culture of biopsy samples have shown positive culture rates in 77–90%.[34,36]

Pleural tuberculosis

A pleural effusion usually occurs 3–6 months after initial infection with *M. TB*. In countries with a high TB prevalence this usually occurs in childhood or adolescence, but it may occur as primary infection later in life. Even without chemotherapy, spontaneous resolution of the effusion almost always occurs.[2] About 50% of cases relapse, however, with a more severe form of TB.[2] Culture of pleural fluid has a sensitivity of up to 35%, usually requiring pleural biopsy for confirmation of the diagnosis of TB.[37–39] Since the diagnostic yield of pleural fluid is low and as pleural biopsy and culture are not always available, diagnostic criteria such as lymphocytic effusion and positive TST have been used.[40] Histology of pleural biopsy tissue samples demonstrates caseous granulomas in up to 60–79% of cases and a diagnosis may be achieved in 86% of cases when combined with microbiological examination.[41,42]

Central nervous system – tuberculous meningitis

Although TB meningitis is a rare disease, it is of importance in view of its significant morbidity and mortality.[43–45] Diagnosis of TB meningitis is usually by cerebrospinal fluid (CSF) examination including microscopy and culture. Microscopy has a low sensitivity, varying from 20 to 58%.[46,47] Culture gives positive results in 50–70% and may take several weeks.[47–49] Diagnostic algorithms have been developed using simple clinical and laboratory data to help in the diagnosis of adults with TB meningitis.[49] If clinical diagnosis is felt to be TB meningitis, samples from other sites for TB cultures should be considered, particularly if there is evidence of disease elsewhere. HIV-infected patients with TB are at increased risk for meningitis, but infection with HIV does not appear to change the clinical manifestations or the outcome of TB meningitis.[50]

Pericardial tuberculosis

TB pericardial disease can present as acute pericarditis, constrictive pericarditis and tamponade. TB pericarditis has a variable clinical presentation and should be considered in the evaluation of a non-resolving pericarditis. Diagnosis is often delayed and development of subacute constrictive pericarditis requiring pericardectomy is common.[51,52] It has been suggested that cardiac tamponade in the early clinical stage of TB pericarditis is the most predictive factor of subsequent constrictive pericarditis.[53] Echocardiography is used as a non-invasive tool to confirm an effusion and to guide fluid aspiration. Diagnosis is usually established by pericardiocentesis with biopsy to obtain pericardial fluid and tissue for mycobacterial culture and histology. Culture of pericardial fluid is usually positive in less than 30% of cases.[33] Positive culture of biopsies has been reported in 70–94%, whereas histology showing granulomatous changes may be seen in 87% of cases.[33,54]

Peritoneal tuberculosis

Abdominal TB has no classical diagnostic symptoms and signs, although abdominal pain and ascites are features usually present on initial clinical presentation.[55–57] Computed tomography (CT) imaging may reveal radiological features of ascites, peritoneal lesions and lymphadenopathy.[58] A diagnosis may be made by positive cultures from ascitic fluid; however, sensitivity has been reported to be less than 10%.[55] Peritoneal biopsy may show histological features of granulomas in 97%[55,59] and positive cultures in 68%.[59] Biopsy may be performed by laparoscopy or laparotomy, but laparoscopy is felt to be safer and superior in the diagnosis of TB peritonitis.[60]

Genito-urinary tuberculosis

Genito-urinary TB may present as pyuria or painless haematuria with sterile urine cultures.[61] If urological TB is suspected, urine cultures should be done as early-morning samples on three consecutive mornings. Microscopy of urine has a low positive microscopy rate and the yield on culture is usually less than 40%.[62,63] Intravenous urography (IVU) is a useful radiological examination, often showing typical features consistent with a diagnosis of genito-urinary TB.[63] Fine-needle aspiration cytology (FNAC) of the kidney may demonstrate granulomatous changes and may give positive culture for AFB. Therefore, it provides a useful means of diagnosing renal TB in patients with negative urine cultures.[64] Renal or bladder biopsy may give a definite diagnosis, if granulomatous changes are found on histology and cultures are positive for *M. TB*. Of patients with pulmonary tuberculosis, 5–8% have positive urine cultures for *M. TB* even though there are no signs, symptoms or laboratory data to suggest genito-urinary tract involvement.[65] In contrast, 21% of patients with extrapulmonary TB have been reported to have positive urine cultures.[66]

Spinal tuberculosis

If undiagnosed, spinal TB can cause spinal cord compression and spinal deformity. Spinal TB may be detected by plain radiographs or by magnetic resonance imaging (MRI) of the spine, which is thought to be useful in early diagnosis of TB spondylitis.[67] Diagnosis can be made by radiologically guided FNAC. Smears may be positive for AFB in 50% of cases and positive cultures have been obtained in 83% of cases.[68] Bacteriological and histological yields have been reported to be similar for surgical biopsy and for percutaneous needle aspiration and biopsy with positive cultures rates of 83%.[69]

Treatment and prevention

Rapid case detection, provision of chemotherapy and ensuring completion of treatment are important in reducing infectivity and transmission. Treatment of TB requires that patients take a combination of drugs [typically quadruple therapy with isoniazid (INH), rifampin (RMP), pyrazinamide and ethambutol or streptomycin] for at least 6 months (12 months for TB meningitis).[5] Failure to adhere fully to the course of treatment may lead to the emergence of drug-resistant strains of bacteria. In 1998 in England and Wales, 6.1% of isolates of *M. TB* were INH resistant and around 1.3% were multidrug resistant.[70] Current internationally accepted empirical therapeutic regimens are likely to be successful in the majority of cases of TB. However increasing rates of drug resistance, especially RMP mono-resistance and multidrug resistance world-wide, mean that drug susceptibility testing is essential for the implementation of optimal therapeutic regimens.

Screening for infected contacts to identify LTBI is currently done using TST and chest radiography. The intention of chemoprophylaxis is to prevent infected individuals from developing active TB. Current recommendations are to give INH for 6–9 months,[71] although many clinicans in the UK give 3 months of INH and rifampicin instead. The risk of developing active disease after infection depends on the bacille Calmette–Guérin (BCG) status, HIV status, whether infection is recent and a number of other factors, including certain co-morbidities and iatrogenic immunosuppression.[70]

The keys to controlling and eradicating TB are fourfold:

1. Case-finding and treating persons with active disease.
2. Tailoring therapy to sensitivity, both to achieve cure in individual patients and to minimise the development of further resistance.
3. Identifying those persons with greatest risk of developing active disease (i.e. LTBI) in the future and providing them with preventive therapy.
4. BCG vaccination in the UK, although some primary care trusts with the lowest prevalences of TB have stopped using it as it provides good protection against disseminated disease in newborns and in infancy but only limited protection against pulmonary disease in adults.

Tests for the detection of active tuberculosis

Given the infectious nature, particularly of pulmonary TB, fast and accurate diagnosis is a very important element of health measures to control the disease.[72] Current strategies are to investigate patients presenting with clinical symptoms using a variety of diagnostic tests, including radiology and microbiology, to establish a diagnosis. Specimen collection is a key element of investigations. For pulmonary TB, respiratory tract specimens are required. Expectorated sputum is thought to be the best specimen, but induced sputum, endotracheal aspiration, bronchial washings or aspirates taken during bronchoscopy, laryngeal swabs and gastric lavage may also be used. Other specimens include cerebrospinal, pericardial, synovial and ascitic fluids and blood, bone marrow, urine and faecal specimens.

Traditional tests for diagnosis of active TB

Microscopy

Microscopy is used to examine clinical specimens or cultures for the presence of AFB. A variety of different stains are available, but the three that are most commonly used are Ziehl–Neelsen, auramine–rhodamine fluorochrome[2,73] and Kinyoun[74] stains. Microscopy indicates that AFB are present in the sample, but does not always indicate viable organisms *per se* or that the organism is *M. TB*. Sputum smears are prepared by spreading purulent portions of the sputum specimen on a glass slide. Approximately 40–50% of patients with pulmonary TB are smear positive[1] (sputum must contain as least 5000 bacilli/ml for them to be detectable by microscopy). It is estimated that 10% of smear-negative patients are also culture negative.[1] Patients with smear-negative, culture-positive TB appear to be responsible for about 17% of TB transmission.[44]

Although microscopy is not very accurate, it remains the most rapid technique and is of value in identifying the most infectious patients for hospital and community infection control. In developing countries, microscopy is often the only test available for diagnosis of pulmonary TB. Sputum-negative cases are usually diagnosed on the basis of clinical and radiological indicators.[75]

Conventional culture-based techniques

Traditionally, mycobacteria have been grown on solid media, containing a cocktail of antimicrobial agents that permit only mycobacteria to replicate. The media are either egg-based [e.g. Lowenstein–Jensen (LJ)[76,77] and Ogawa media[78]] or agar-based (Middlebrook 7H9, 7H10 and 7H11[77]). Media are assessed for rate and appearance of growth of mycobacteria. The growth of *M. TB* colonies is distinct, producing either beige-coloured, rough, dry, corded, flat colonies with irregular borders or warty, granular colonies that with time heap into a cauliflower shape.[79]

Culture is the most sensitive of currently available tests (sensitivity rates of up to 98% have been reported), and also permits identification and drug sensitivity tests to be made. However, it may require up to 6–8 weeks for the isolation of *M. TB* from a clinical specimen and in 10–20% of cases the bacillus is not successfully cultured.[72] When material from solid cultures is used, *M. TB* can often be distinguished from atypical mycobacteria.[80] Culture is more expensive than microscopy and requires a high standard of technical competence.

Serological tests

Numerous serological tests for TB have been developed over the years using a variety of antigens to detect certain antibodies in the blood, including complement fixation tests, haemagglutination tests, radioimmunoassays and enzyme-linked immunosorbent assays (ELISAs).[81] They have been extensively evaluated in developing countries, since they are less expensive, rapid and simple, making them ideal for use in resource-poor settings.[82] It has been suggested that serological tests might be useful in combination with other tests in diagnosis of smear-negative TB in settings, where culture is not routinely available.[83] However, so far none of these tests have shown adequate accuracy, so they have not been widely implemented. Antibody responses are directed against a broad set of antigens, responses vary individually and sensitivities have generally been very poor.[72,84] Sensitivities of 16–57% and specificities of 62–100% have been reported.[85] It has been estimated that even 30% of patients with smear-positive pulmonary TB do not have detectable antibody to any single reagent.[81] Since exposure to atypical mycobacteria, BCG vaccination and HIV prevalence influences results of serological tests, accuracy reports of these tests vary in different settings.[2]

Adenosine deaminase and cytokine assays

Several biochemical markers have been investigated for their potential in diagnosis of TB. An enzyme produced by lymphocytes, adenosine deaminase (ADA), has been studied mainly in the diagnosis of pleural TB, peritoneal TB and TB meningitis. Several cytokines have also been evaluated for the diagnosis of TB, including interferon-γ and TNF-α.

As diagnostic tests, ADA and interferon-γ assays offer several advantages, they are rapid, simple, non-invasive (especially in diagnosis of pleural TB) and can be performed in most clinical laboratories.[86] However both tests are thought to lack accuracy if used alone. For example, estimation of ADA and interferon-γ can be useful in differentiating malignancy from tuberculous pleural effusions, but other conditions such as empyemas are less easy to differentiate.[87]

New tests for diagnosis of active TB

The traditional mainstay and gold standard of diagnosis is the combination of rapid identification of bacilli on direct microscopy combined with culture for subtype and antibiotic sensitivity. However, the diagnostic limitations of microscopy and the length of time required for traditional culture methods have focused attention on developing rapid methods for the detection of *M. TB* in clinical specimens and the early identification of mycobacterial isolates.

Rapid liquid culture systems

With traditional culture methods such as the use of LJ media, the time to detection of mycobacterial growth may be 4–6 weeks.[88,89] Faster culture of mycobacterial isolates has been achieved with manual culture systems [Septi-Chek AFB or the manual mycobacterial growth indicator tube (MGIT)].[90] Compared with automated systems, however, manual systems have clear disadvantages, including a longer time to detection of significant mycobacteria, more technical hands-on time and a higher contamination rate.[90]

The time to detection of growth of a mycobacterial species can be shortened significantly with the use

of automated or semi-automated liquid culture systems. Radiometric liquid culture using a broth of radiolabelled carbon has been used for many years. However, because of increasing problems with handling and disposal of radioactive waste, use of expensive media and staff time, systems that rely on non-radiometric growth have been developed.[91] Nevertheless the semi-automated radiometric culture BACTEC 460TB system remains the fastest (14–17 days) and is widely accepted as a reference standard.[89]

Several non-radiometric automated or semi-automated liquid culture systems have been introduced, including the MB/BacT or MycoBacT system, BACTEC 9000MB, Bactec MGIT 960 and ESP Myco and Accumed/Difco ESPII. These systems measure changes in gas pressure, carbon dioxide production or oxygen consumption fluorimetrically or colorimetrically.[92] They allow continuous monitoring of cultures and there is no need for further operator input after loading the specimen.[92] Multiple studies have been performed comparing different media systems. Major parameters in comparisons between these systems are recovery rates for mycobacteria, time to detection and contamination rates. Optimal recovery is usually achieved through a combination of rapid automated liquid culture systems and solid media.[77] Nevertheless three sputa still seem to be required for accurate diagnosis of TB.[93]

M. TB often exhibits a characteristic morphological pattern (serpentine cording) when grown in liquid media. This has been used for rapid presumptive identification of *M. TB* and other mycobacterial species.[74,94] The radiometric Bactec 460 and also the automated non-radiometric liquid culture systems also allow susceptibility testing.[95–98] Reductions in turnaround times from 21 days for LJ-based tests to 6–12 days for automated liquid culture systems for susceptibility testing have been reported.[98,99] The liquid culture methods are expensive, however, and require elaborate technology.[100] In addition, skilled and experienced staff are crucial as contamination rates have been high when inexperienced and untrained staff have used these systems.[92]

Methods for rapid detection and identification directly in clinical specimens

The slow growth of *M. TB* has led to the development of methods to detect *M. TB* directly in clinical specimens.

Nucleic acid amplification tests

Nucleic acid amplification tests (NAATs) are molecular systems which are able to detect small amounts of genetic material (DNA or RNA target sequences) from the micro-organism, and based on repetitive amplication of target sequences. If the target organism is not present in the sample, no amplification will occur. A variety of amplification methods may be used, including amplification of the target nucleic acid, such as the polymerase chain reaction (PCR), or amplification of a nucleic acid probe, such as a ligase chain reaction.

PCR is the most common of these methods. The products from the PCR reaction are usually analysed on an agarose gel, which separates the DNA products according to size against a molecular weight marker. This determines whether the DNA between the two primers in a particular strain is of the expected size. Detection of the amplified products can also be done by DNA sequencing, an enzyme immunoassay format using probe-based colorimetric detection or by fluorescence emission technology.[101] Molecular amplification with a probe is usually a two-step process requiring the amplification of the DNA region of interest and hybridisation of a DNA probe to a specific element such as the insertion sequence IS6110. A hybridisation signal, normally a colorimetric or fluorescent signal, is then detected. The genetic material can be used to identify species and can sometimes be used to identify the genes that code antibiotic resistance.

Although the specificity of a well-designed PCR can be high, the sensitivity is thought to be less than that of culture, but can be optimised by performing PCR on high-quality specimens. The sensitivity of a carefully performed quality-controlled PCR would be expected to be 90–100% smear positive and 60–70% on smear-negative, culture-positive sputum samples.[92]

Commercially available tests include the Roche Amplicor® *Mycobacterium tuberculosis* test[102] (PCR target amplification of part of the 16S rRNA gene, followed by colorimetric detection of the PCR product). The Gen-Probe Amplified *Mycobacterium tuberculosi*s Direct Test (MTD®),[103] which is an isothermal transcription amplification method, uses rRNA as the target rather than DNA. The BD ProbeTec (multiplex strand displacement system) also uses an isothermal amplification but with a DNA fragment.[104]

In-house PCR tests have been developed owing to the high cost of commercial tests. The majority of

these tests are based on the IS6110 insertion sequence owing to its presence in multiple copies in most isolates. Using this technique as a diagnostic test can lead to PCR false positives as mycobacteria other than TB contain this insertion element,[105] and to false negatives[106] as not all *M. TB* isolates contain a copy of the IS6110 element. In addition to IS6110, other target genes include MBP64, rpob and hsp65.

PCR reduces the time for identification of *M. TB* and may be completed within 3–6 hours after the receipt of the specimen.[102] PCR is not used routinely, especially in developing countries, in view of considerable cost and laboratory equipment and skills required.[107] However, PCR can be carried out on a crude extract directly from clinical samples; therefore, in resource-poor settings where culturing is difficult, PCR is often seen as an attractive alternative.[108] Since PCR is able to detect small numbers of organisms, cross-contamination in busy clinical laboratories might yield significant numbers of false-positive results.[109]

Ligase chain reaction (LCx® Tb test) is based on the amplification of a segment of the chromosomal gene of *M. TB* encoding for the protein antigen b. This gene sequence appears to be specific of the *M. TB* complex and has been detected in all *M. TB* complex strains examined to date.[110] High sensitivity and specificity of the test have been reported;[111] however the LCx® test has recently been withdrawn from the market owing to batch problems.

Mycobacteriophage-based methods

Mycobacteriophage tests are an alternative to PCR tests and may be useful for resource-poor countries where PCR is impractical. Rapid phenotypic-based methods have been applied directly to clinical specimens, although greater success has been achieved with cultured isolates.[80] These tests have the advantage of being easy to perform and present a low-cost means to screen for antimicrobial resistance.[112] Mycobacterial cultures are infected with mycobacteriophage and exogenous, non-infecting phage are killed. The signal is amplified biologically by replication of the phage within mycobacteria and detected normally by one of two methods. The simplest of these is the phage amplified biological (PhaB) assay, where the phage is plated on to a lawn of the rapidly growing *M. smegmatis*, which is also lysed by the phage and a numerical result is obtained relating to the number of viable mycobacteria in the original sample.[112,113] Alternatively, a luciferase reporter phage may be used. When infecting viable mycobacteria, it produces quantifiable light that is not observed if drug-sensitive mycobacteria are rendered non-viable by treatment with active antimicrobials.[100] Light can be detected with a Polaroid film box.[114] Either of these methods can also be used to determine drug resistance by incubating the culture with the relevant antibiotic, as only viable mycobacteria will be detected by the phage assay.

Methods for the rapid identification of mycobacterial species from cultured isolates and drug susceptibility testing

High-performance liquid chromatography (HPLC)

HPLC is used to analyse mycolic acids extracted from an unknown organism with ultraviolet or fluorescence detection: the HPLC pattern is compared with a library of known patterns, usually facilitated by a decision analysis system. HPLC methods have shown high sensitivity and specificity.[92] However, high equipment costs and the level of expertise required for the analysis have restricted its use.[115]

Bactec NAP test

NAP-selective inhibition of *M. TB* complex is a conventional biochemical test used in species identification that has been adapted for use with the radiometric Bactec system.[116] *M. TB* and *M. bovis* are both susceptible to NAP (a chloramphenicol-related compound that inhibits growth), whereas atypical mycobacteria are resistant to it.[117] The Bactec radiometric growth system can be used for rapid presumptive identification of *M. TB* where NAP is used along with a growth control tube.

Nucleic acid probes

A nucleic acid probe, such as the AccuProbe *M. TB,* has significantly reduced the time required to isolate and identify *M. TB*.[30,118] The AccuProbe system can be used for the identification of *M. TB* complex, *M. avium complex*, *M. avium*, *M. intracellulare*, *M. kansasii* and *M. gordonae* from culture.[119] The bacterial membranes are lysed, releasing the DNA into solution. DNA probes complementary to the bacterial target sequence, specific to the organism being identified, are used to identify whether a culture, once grown, is *M. TB* or atypical. The AccuProbe is rapid and simple to perform and takes about 1–2 hours from culture.[120] Although certain of the more clinically significant atypical mycobacteria can be identified, such as *M. avium complex*, a separate test must be performed for each species which is tested for.[121]

Other rapid identification systems include the LIPA mycobacteria kit, which is a PCR-based reverse hybridisation line probe assay. The Inno-LIPA Rif TB identifies 95% of RMP-resistant isolates.[122] Both Accuprobe and LIPA mycobacteria tests have been reported to have accuracies above 90%, but the cost of such probes limits their routine diagnostic use.[123,124]

Nucleic acid amplication-based methods and DNA sequencing
In developed countries, PCR-based assays are often routinely used to identify cultures such as the Hain DNA strip for the identification of *M. TB* complex and non-tuberculous mycobacteria (NTM) strains[125] or the INNO-LIPA assay for *M. TB* identification and rifampicin resistance.[122] Direct detection of *M. TB* in respiratory specimens by DNA sequencing is used to identify drug resistance.[126] RMP resistance is important, since it is often a surrogate for MDR-TB.[127] About 95% of all *M. TB* RMP-resistant clinical isolates harbour specific mutations within a region of the rpoB gene.[128,129] In contrast to RMP, genotypic testing for INH resistance is much more complex and alterations in several genes including *katG* and *inhA* have been reported.[129] This technique uses high-cost equipment beyond the reach of most clinical laboratories except at the reference level.

PCR restriction enzyme analysis
PCR restriction enzyme analysis (PRA) is based on the PCR amplification of a fragment of the *hsp65* gene, followed by restriction.[130] It is a simple and rapid identification method and the turnaround time may be 24–48 hours.[131,132] PRA seems to be an efficient method for the identification of mycobacteria to the species level.[133] High accuracy of this method has been reported at relatively low cost compared with, for example, identification of *M. TB* isolates with the Accuprobe.[132]

Mycobacteriophage-based methods
See the section above.

Current recommendations for rapid diagnosis

In the UK, the Public Health Laboratory Service Mycobacterium Reference Unit has proposed a streamlined model to speed up diagnosis in smear-positive patients in London and comparable cities.[134] In the normal course of diagnosis, hospitals perform microscopy and culture (usually on to solid media but increasingly using liquid culture media); over 95% of hospitals then refer these cultures to the reference unit for identification and drug susceptibility testing. In the more streamlined model, hospitals send AFB-positive sputum samples directly to the reference laboratory for rapid culture. Cultures positive for AFB would then be analysed using DNA hybridisation probes or PCR for the identification of *M. TB* or *M. avium complex*. *M. TB*-positive specimens would then be inoculated into a rapid culture system for INH and RMP detection and on to solid media for detecting resistance to other first-line drugs. The whole process should be completed with 30 days. For patients with a high clinical risk of resistant TB, molecular amplification methods would be used to identify TB and predict RMP resistance and these results would be available in 3–4 days, from taking primary specimens.

Tests for the detection of LTBI

Available tests for the detection of LTBI are as follows.

Tuberculin skin test

TST is currently the standard tool to detect latent TB infection, although it is far from a 'gold' standard. TST is based on the detection of delayed-type hypersensitivity to purified protein derivative (PPD), a mixture of antigens shared by several mycobacteria that gives rise to a skin reaction. Two visits are required for the test, one for PPD inoculation (the Mantoux technique uses intracutaneous injection by needle and syringe) and another after 48–72 hours for interpretation of the result based on the size of the skin reaction.

TST is relatively cheap and can be performed without the need for a specialist laboratory. Difficulties in test administration and interpretation often lead to false results. There are many practical difficulties in conducting TST. The second of the two visits might pose a compliance problem for people who live in remote settings and in some patient groups; for example, in urban HIV clinics return rates are low.[135] The inoculation induces painful skin inflammation sometimes with induration[136] and scarring at the injection site, which may be unacceptable to certain population groups. The test might not be possible in individuals with skin disorders.

Dose of PPD, method of application and criteria for interpretation vary between countries. Weak PPD doses increase the likelihood of false-negative results and strong doses increase the likelihood of false-positive results. A 1.5-mm difference of reaction size may be seen when a 10-tuberculin unit (TU) dose is compared with a 5-TU dose.[137]

The technique for inoculating PPD doses may cause false results.[29] For example, the Heaf test, used in the UK, is usually less precise than the Mantoux test, although the two tests generally correlate.[70] Different cut-offs are used for positivity of TST, as there is no general consensus on this issue. Criteria of 5, 10 or 15 mm for skin reaction have been recommended depending on the clinical situation.[138] There can be false TST results from operator variability in both inoculation and reading of the test.[139] Digit preference, for example rounding measures of TST induration to the nearest multiple of 5 mm and interpretation bias, can significantly affect TST results.[140]

There are many reasons for false-positive TST results. PPD contains a poorly defined mixture of mycobacterial antigens. Because antigens are shared with other mycobacteria, tuberculin reactivity leading to a positive TST can result from BCG vaccination with a live attenuated mycobacterial strain derived from *M. bovis* or from exposure to atypical mycobacteria.[141–144] The effect of BCG vaccination on TST can persist as long as 15 years after vaccination.[145] Reaction due to BCG vaccination tends to be small, but this is not always consistent.[138] Specificity problems of PPD can be addressed by simultaneous skin testing with *M. TB* PPD and sensitins, which are PPD-like products derived from atypical mycobacteria. This approach can help to discriminate patients with TB from those who are infected with *M. avium complex*.[72,146,147] Repeated TSTs may induce booster responses leading to false-positive results.[148] Anergy associated with HIV infection, disseminated TB or immunosuppression due to haemodialysis, transplantation or medication can give rise to false-negative reactions.[29,41,149]

There have been many responses to the above problems, none of which are ideal. Some countries have stopped using either BCG vaccination altogether, or school-age vaccination, as BCG has only limited effect on preventing adult tuberculosis. An alternative strategy (and the policy of the USA and The Netherlands) has been to use TST to identify recently infected individuals and give isoniazid chemoprophylaxis. The British Thoracic Society no longer recommends performing TST among BCG-vaccinated people with recent TB exposure.[70]

Immune-based blood tests for the rapid identification of latent tuberculosis infection: the interferon-γ assays

Interferon-γ assays have been developed as tests to replace TST. In this scenario, blood samples would be taken from the patient and incubated with mycobacterial antigens specific for *M. TB* complex strains and absent from the BCG vaccine strain. T lymphocytes within the blood sample produce interferon-γ as a marker of infection or active TB.[32] Since *M. TB* is an intracellular pathogen, assessment of whether a patient's T cells have been exposed to and sensitised by antigens specific to *M. TB*, may provide an alternative approach to diagnosis.[72]

The antigens used to elicit an interferon-γ response define the two main types of the available tests: assays based on PPD and those based on RD1-specific antigens including early secretory antigen target 6 (ESAT-6) and culture filtrate protein 10 (CFP-10). Various commercial and in-house tests based on PPD, ESAT-6 and/or CFP-10 have been evaluated using either an enzyme-linked immunospot assay (ELISPOT) or an enzyme-linked immunosorbent assay (ELISA). The two commercial tests using ELISA are Quantiferon® (based on PPD) and Quantiferon Gold®[150] (based on ESAT-6 and CFP-10). The T SPOT-TB® assay[151,152] is an ELISPOT assay and is also based on RD1-specific antigens.

Interferon-γ assays have several advantages over TST. They involve having a blood test at a single visit and a return visit might not be needed in some settings, depending on the test result. Automated testing has the advantage of reducing reader bias as interpretation is objective. A booster phenomenon does not occur and therefore screening of people who are repeatedly exposed to TB (e.g. healthcare workers) becomes feasible. Interferon-γ assays might improve diagnostic accuracy in latently infected people with greatest risk of progression in whom TST is often false negative, e.g. people with HIV infection.[153] Interferon-γ assays also have some limitations. The need to perform a blood test might not be desirable to certain patient groups. The blood often needs to be processed within 12 hours after collection and laboratories need to gain expertise in technology like isolation of mononuclear cells.[41]

The Quantiferon test measures interferon-γ production after *in vitro* stimulation of whole blood cells with PPD from *M. TB* and control antigens.[154,155] It is able to discriminate between *M. TB* and *M. avium intracellulare complex* infection.[156] It responds to multiple antigens spontaneously.[154,157–159] It does not boost anamnestic immune responses.[159] Two key disadvantages of Quantiferon are that it can give false-positive results in BCG-vaccinated people

and that it does not discriminate between most of the atypical mycobacteria and *M. TB*.[53,160,161]

Interferon-γ assay based on RD1-specific antigens, ESAT-6 and CFP-10, can overcome some of the above disadvantages. Comparative genomics has identified several genetic regions in *M. TB* and *M. bovis* that are deleted in all tested BCG strains.[162] The identified region, so called RD1 region, is present in *M. kansasii*, *M. szulgai* and *M. marinum*.[163,164] Proteins encoded in these regions have formed the basis of new specific T cell-based blood tests that do not cross-react with BCG, but only two antigens, ESAT-6 and CFP 10, have been studied in detail in humans.[165] ESAT-6 is a secreted antigen that is expressed in the *M. TB* complex, but is absent from BCG and most atypical mycobacteria.[163,166] ESAT-6 and CFP-10 share the same messenger RNA transcript, which suggest that they may interact with one another and serve a common function in detection of *M. TB*.[167] All stimulated T-lymphocytes secrete interferon-γ, but the ESAT-6- and CFP-10-specific assays can only detect interferon-γ secreted from T-lymphocytes produced as a result of exposure to ESAT-6 and CFP-10 antigen. *In vivo* and *in vitro* experiments have shown that the combination of ESAT-6 and CFP-10 has a higher sensitivity and specificity than PPD in diagnosis of TB infection.[53,160,161]

Use of interferon-gamma assays for active TB infection

Interferon-γ assays are being assessed for use in people with suspected active TB infection.[168] The key problem with their use in this context is that they detect LTBI and not only active disease.[41,72] Where the prevalence of TB is low and clinical indications are strong, a positive test result may be assumed to indicate active TB. However, in TB-endemic countries a positive test result would be less meaningful, potentially indicating only latent TB infection.[162] However, since *M. TB* infection is a necessary prerequisite for active TB disease, a negative test result can effectively exclude a diagnosis of TB, if the test has sufficiently high diagnostic sensitivity. These tests could potentially serve as useful 'rule-out' tests in patients with suspected TB in low and high prevalence countries.

Problems in the assessment of diagnostic tests for TB

The main sources of bias that may be relevant in the assessment of diagnostic test studies can be broadly categorised as relating to the study population, the selection and execution of the tests, interpretation of the tests and the data analysis and presentation and in practice with the quality of the samples submitted. In general, studies evaluating diagnostic tests fail to address these issues adequately, although this may in part be due to poor reporting. In a methodological review, only one out of seven quality standards (the avoidance of verification bias) was fulfilled by more than 50% of the 112 eligible studies retrieved.[169] Empirical evidence for the impact of many of these quality features on test accuracy is still limited. Two studies[170,171] found several features that significantly over- or underestimated test accuracy, including the use of case–control design with healthy controls and severe cases of disease, use of different reference tests, selective inclusion of patients and retrospective data collection.[171]

Errors in the design or reporting of TB diagnostic studies are said to be particularly common,[172] including failure to describe methods for selection and enrolment of patients, inadequate sample size and declaration of positive and negative predictive values even when the test population in no way resembles the population for which the test is intended. Often inadequate gold standards for clinical case definition and microbiology are employed. Failure to state a specific research question or test indication under study and routine over-statement of the significance of trial results or implication of clinical impact not supported by data can be found in studies assessing diagnostic tests for TB. Furthermore, analyses are often performed on a 'per sample' basis such that patients from whom several samples are taken will affect results.[102] *Post hoc* analyses may be conducted restricting each patient to no more than three samples, or by using 'per patient' analyses, but interpretation of results is limited by the need to consider how to resolve situations in which test results differ in different samples from the same patient.

Use of an appropriate reference test

Standard techniques for assessing diagnostic tests assume that a definitive reference test is available, that is, that the reference test used is as close to 100% accurate as it can be. However, it may be either that the available test is far from perfect, or that such a test simply does not exist. The standard reference test for detection of TB infection has historically been culture alone. However, culture may fail to detect mycobacteria that can be picked up by, for example, nucleic acid

amplification tests, and will therefore incorrectly classify patients with TB as false-positive results.[173] Serious inaccuracies in the reference test will lead to over- or underestimation of the true accuracy of a new test, in this case where the index and reference test are not conditionally independent, that is, may make the same errors, the accuracy of the new test will be overestimated,[174] potentially appearing perfectly accurate regardless of its association with true disease status.[175] Although there is a role for studies demonstrating 'proof of principle' for new diagnostic systems, the best way to evaluate how new tests will perform in practice is to adopt a reference strategy, where the reference diagnosis is made on the basis of clinical information in combination with a battery of other tests.[174] It should be borne in mind, however, that no symptoms are pathognomonic for TB. Although it is true that for pulmonary TB, clinical symptoms and radiology may have a high specificity and sensitivity for diagnosis of TB, this is not the case for extrapulmonary TB where the accuracy of clinical diagnosis is more limited and laboratory culture/histopathology provides definitive diagnosis.

The situation for LTBI is even more complex because the TST was, until recently, the only test available to detect LTBI. It has therefore been used as the reference standard world-wide although it has widely recognised limitations as a gold standard (as described above). Long-term follow-up to compare the number of incident cases in healthy, TB-exposed individuals who are positive by TST or positive by the new interferon-γ tests, could potentially definitively determine the accuracy of both tests. However, this approach would require extremely large numbers of TB-exposed subjects and many years of long-term follow-up, would be very expensive and would be limited by the fact that subsequent exposure to other active TB cases is impossible to control for. One potential solution, and the one adopted for our analysis in Chapter 15, is to evaluate the tests indirectly. Given that the risk of TB infection is greatest among those contacts who share a room with the index case for the greatest length of time, that is, airborne transmission increases with length of exposure and proximity to an infectious TB case,[21,176–178] it follows that accuracy of tests for latent TB infection should be associated with level of exposure. This can be evaluated in observational studies that have ascertained approximate TB exposure in a relevant population and setting (e.g. outbreak investigations), performed various index tests of interest in all eligible subjects and compared test results with exposure status. It should be noted that length of exposure alone is not the whole story – it is a surrogate (albeit a good one) for the two critical factors: the output of bacilli from a patient and the length of unprotected exposure to that output (which is usually greatest for household or 'household equivalent' or healthcare workers). Studies have demonstrated apparently high levels of infection from very short exposures to highly infectious cases.[44,179,180] This potential confounding variable does not affect results from institutional outbreaks arising from a point source exposure (i.e. a single infectious source case), but may affect results from community-based studies where different contacts have been exposed to different infectious source cases, unless the infectiousness of each source case has been individually assessed and incorporated into the calculation of amount of TB exposure.

Blinding

The interpretation of many diagnostic tests involves some degree of subjective interpretation. In clinical practice, test interpretation can be influenced by both the knowledge of the results of other tests and the specific clinical characteristics of the person being tested. To avoid bias in the evaluation of diagnostic tests, a 'blinded' study should be performed, where both tests are interpreted without knowledge of the clinical characteristics or the test results.[181] This ensures that it is only the diagnostic contribution of the test itself that is being evaluated.

Study design

Cohort studies assemble patients at risk for a disease in whom both the new test and the reference test are performed, whereas case–control studies assemble patients with the disease and controls without the disease (on the basis of the reference test results) and compare the index test results in both groups.[182] Case–control studies tend to be at higher risk from bias: cases tend to be selected on the basis of a positive reference test result and the result of the test under evaluation ascertained after true disease status is known; the prevalence of the target disorder tends to be higher than in cohort studies (or than in practice); and cases and controls are often selected from opposite ends of the disease spectrum, for example, severe cases and healthy controls.[183] The 'best' cohort studies are prospective in design, with consecutive recruitment of patients; this allows evaluation on the full spectrum (see below) presenting in that setting, the collection of appropriate baseline information and implementation of rigorous protocols for testing.

Clinical heterogeneity

There is some limited evidence that test accuracy statistics may not be generalisable from diagnostic test studies to patients in clinical practice as a result of variations in case mix of participants. The term 'spectrum' refers to the range of pathological, clinical and co-morbid patient or disease characteristics in a study sample and 'spectrum bias' has been used to describe scenarios where the accuracy indices obtained in one study cannot be assumed to apply to other patients in other contexts and also where test accuracy has been seen to vary according to subgroups of patients within the same study. Such characteristics can be likened to effect modifiers in therapeutic interventions.

It is often assumed that indices of test accuracy such as sensitivity and specificity are fixed (for any given threshold) and that what varies is the predictive value between groups with different disease prevalence, the effect of which is easy to estimate. However, theoretical examples[174,182,184] indicate that where spectrum bias is present, either sensitivity or specificity would be expected to change. Variations in case mix, therefore, may affect the generalisability of a study's accuracy results.

Chapter 2

Research questions addressed

Aims

We aimed to evaluate the effectiveness of available rapid diagnostic tests to identify TB infection.

The detection of active tuberculosis infection

For patients presenting with suspected active TB infection, of any form, the following questions were addressed:

- How accurate are nucleic acid amplification tests at detecting *M. TB* in clinical samples, and are the commercial versions superior to 'inhouse' versions of the tests?
- How accurate are molecular probe tests at detecting *M. TB* in clinical samples?
- How accurate are serodiagnostic and biochemical tests at diagnosing TB infection?
- How accurate are phage-based tests at detecting *M. TB* in clinical samples?
- What is the value of fully automated liquid culture systems over and above standard culture using either solid or liquid media?
- What is the most cost-effective way of using these tests to diagnose active TB?

The detection of latent tuberculosis infection

For patients with potential LTBI, the following question was addressed:

- What is the value of interferon-γ assay tests over and above the TST?

Objectives

The objectives were as follows:

1. For each form of active tuberculosis:
   (a) to conduct systematic reviews to evaluate the accuracy of nucleic acid amplification tests in patients presenting with clinical signs and symptoms
   (b) to conduct systematic reviews to evaluate the accuracy of molecular probe tests in patients presenting with clinical signs and symptoms
   (c) to conduct systematic reviews to evaluate the accuracy of serodiagnostic tests and other biochemical tests to detect impaired immunity in patients presenting with clinical signs and symptoms
   (d) to conduct systematic reviews to evaluate the accuracy of phage-based tests in patients presenting with clinical signs and symptoms.
2. To conduct a systematic review to evaluate how effective fully automated liquid culture systems for diagnosing active TB are over and above standard culture.
3. To conduct a systematic review to evaluate the use of interferon-γ assays for detection of LTBI.
4. To examine the likely NHS and societal consequences of the false-positive and false-negative rates resulting from use of the key tests for active pulmonary TB or LTBI.

Chapter 3
Methods

Systematic reviews of diagnostic tests for detection of active tuberculosis infection

Inclusion criteria

Population

Studies of adults or children with any form of active TB were eligible for inclusion. Patients with any co-morbidity (including HIV infection) were included. Studies exclusively conducted in patients with non-tuberculous mycobacterial infection were excluded on the basis that these infections are rare and inclusion of them was outwith the resource constraints of the review.

For most tests, studies with more than one specimen per patient were included only where accuracy data could be extracted on a per patient as opposed to a per specimen basis or where the difference in the number of specimens compared with the number of patients was less than 10%. For the fully automated liquid culture tests, all relevant studies were included as none of the studies provided per patient data, and given the widespread adoption of this new and expensive test, we judged it to be important to include this group of studies. Studies of specimens 'spiked' with mycobacteria were excluded as they did not use clinical samples.

Diagnostic tests

Any study that compared a **rapid** test for detection of active TB with a reference standard was included. 'Rapid' tests were defined as those tests for which a result could be obtained in less than the time taken for standard culture (on solid or liquid media). Tests eligible for inclusion were as follows:

1. fully automated liquid culture techniques
2. all methods for the rapid identification of *M. TB* directly in clinical specimens
   (a) nucleic acid amplification tests
   (b) molecular probe tests
   (c) phage-based tests
3. 'serodiagnostic' tests
   (a) anti-TB antibody tests
4. other biochemical tests reflecting the local immune response to *M. TB*
   (a) ADA
   (b) antigen tests
   (c) cytokine tests
   (d) lysozyme tests
   (e) miscellaneous other tests.

Studies evaluating tests used for strain typing of TB were excluded, as these are more of an epidemiological tool than tests for use in routine clinical practice. Studies evaluating drug susceptibility tests were also excluded, as they were beyond the scope of this project, which was focused on tests assisting the primary diagnosis of active TB and LTBI.

Reference standards

Reference standards for tests for detecting active TB can be defined as follows:

A: culture and/or microscopy smear test
B: very high clinical suspicion of TB ± response to therapy
C: clinical suspicion of TB, but it is not certain one way or the other.

Studies may use one or more of these reference tests either alone or in combination with each other as a reference strategy. Strategy A alone, although previously considered good practice, is now recognised as an inadequate reference standard, especially in patients with AFB smear-negative tuberculosis. Although culture specificity is high (a positive culture result is highly indicative of the presence of mycobacteria), sensitivity is much poorer as culture can miss true cases of TB. Unfortunately, clinical diagnosis, although improving sensitivity, has a relatively low specificity for TB diagnosis, particularly for extrapulmonary TB. The definition of strategies B and C can also vary significantly, in terms of what signs and symptoms are considered to suggest the presence of TB infection. We accepted any of these categories as eligible reference tests and examined any impact on accuracy in the analyses by designating culture plus high clinical suspicion with or without additional investigations as an ideal reference strategy. However, we recognise that for extrapulmonary TB this situation is less clear cut and there is no obvious gold standard.

Study setting

No restrictions on study setting were applied and studies from all countries were eligible for inclusion.

Study design

Only 'cohort' or case series type studies that compared a diagnostic test with an established reference standard in patients **suspected of having** tuberculosis were eligible for inclusion in the review. These could be either prospective or retrospective in nature.

'Case–control'-type studies where the performance of a test is compared in two or more groups of patients potentially ranging from those with confirmed active TB infection through to those with diseases other than TB or even no known disease (healthy controls) were excluded. This type of design is known to be significantly more susceptible to bias than cohort studies, especially when healthy control patients are included, the artificial selection of patients leading to an unrepresentative case mix of patients.

Outcome measures

The evaluation of diagnostic tests has largely focused on the establishment of test accuracy, and this was the main focus of this review. Studies that examined the effect of diagnosis on diagnostic thinking, patient management or subsequent patient outcomes were also eligible for inclusion, but none were identified. Studies focusing on the establishment of technical efficacy alone were excluded.

At a minimum, we required accuracy studies to report sufficient information to allow the construction of a 2×2 contingency table. This information was used to calculate relevant accuracy statistics. Studies reporting only summary accuracy statistics without sufficient raw data to allow the construction of a 2×2 table were excluded. For studies using discrepant analysis (where false-positive and/or false-negative results usually against culture are resolved by examining clinical data for those patients), pre-discrepant analysis results were used wherever possible, as this can be a potential source of bias.[185]

To account for varying definitions for an abnormal result between studies, data were extracted at a variety of cut-off points where possible. In general, only one data set per test comparison was included in each analysis; where possible, the study authors' recommended cut-off was used, otherwise the cut-off that appeared to give the best result in terms of joint sensitivity and specificity was selected.

A *post hoc* amendment was made to this section following screening of the fully automated liquid culture (FALC) studies. None of the identified studies provided 2×2 data for the FALC studies against a reference test for groups of patients **suspected** of having TB. Instead, specimens known to be infected with mycobacteria (identified via PCR and other biochemical tests) were cultured using FALC plus at least one other standard culture method (liquid or solid media). Therefore, only sensitivity data could be provided along with time to detection and contamination rates.

Literature search

Literature was identified from several sources, including electronic databases and other sources (see Appendix 1 for a detailed list). A comprehensive database of relevant articles was constructed using Reference Manager. All databases were searched from 1975 to August 2003. Reference lists of included studies and relevant review articles were scanned to check for additional studies not identified from other sources.

In the first instance, searches were not restricted to English language only as, in principle, all eligible studies should be included in a systematic review regardless of language of publication. However, owing to the volume of non-English language literature identified and restrictions on translation due to time-frame and resource constraints, we were ultimately unable to assess these studies for inclusion.

A highly sensitive strategy to identify studies of tests evaluated in patients with active TB infection was used. Owing to the high volume of studies in TB infection, we opted to combine tuberculosis-related terms first with terms relating to the tests under evaluation and second in combination with a sensitive methodological filter developed to identify diagnostic accuracy studies (Appendix 1).

Study inclusion

Studies were selected for inclusion in the review in a two-stage process. In the first instance, the literature search results (titles and abstracts) were screened independently by two reviewers to identify all citations that appeared to meet our inclusion criteria as described in the section 'Inclusion criteria' (p. 15). Full manuscripts of all selected citations were retrieved. Where it was not possible to determine study eligibility from the title and/or abstract, the full manuscript was obtained. A checklist for study inclusion was piloted and

subsequently completed for every full paper retrieved (see Appendix 2). Any disagreements over study inclusion were resolved by consensus or, if necessary, by arbitration by a third reviewer.

Quality assessment

The methodological quality of all included studies was appraised using a formal quality assessment tool developed by the University of York (also funded by the HTA programme).[51]

Use of a formal quality assessment tool allows the exploration of study design aspects either for which empirical evidence of bias exists[170,171] or that are generally accepted as important for diagnostic test studies. Appendix 3 provides a list of quality assessment criteria used and a guide to their interpretation.

Study quality was assessed independently by two reviewers. Any disagreements were resolved by consensus or, if necessary, by arbitration by a third reviewer.

Data extraction

The extraction of study findings was conducted in duplicate using a predesigned and piloted data extraction form to minimise any errors. Data were recorded on to a Microsoft Access database. Information on study participants, study design, tests and reference test details, test performance (2×2 contingency tables) and potential sources of bias were extracted (the full data extraction form is provided in Appendix 4) . Any disagreements between reviewers were resolved by consensus or, if necessary, by arbitration by a third reviewer.

Data synthesis

In the first instance, studies were grouped according to:

1. type of TB infection
2. type of test (rapid culture, nucleic acid amplification in clinical specimen, etc.).

Separate reviews were undertaken for each combination of the above factors. Individual tests within a group of tests were evaluated where sufficient studies are available. Analyses were performed using STATA version 8.

For each test comparison, the sensitivity, specificity and their exact 95% confidence intervals (CIs) were calculated. Statistical heterogeneity of sensitivities and specificities was assessed using the χ^2 test and by plotting sensitivity against the false-positive rate (one minus specificity) on a receiver operating characteristic (ROC) plot and visually considering the scatter of points.

Sensitivity and specificity are not independent of each other as they vary with the threshold for test positivity. As it is likely that explicit differences in threshold occurred between some studies when different cut-points were used to define positives, and implicit differences may have occurred owing to differences in case mix, we did not carry out direct pooling of sensitivity and specificity or likelihood ratios, but undertook an analysis of summary receiver operating characteristic (SROC) curves which allow for variation in threshold. The method is based on estimating a single summary measure of test performance per study known as the diagnostic odds ratio (DOR) (a statistic which describes the ratio of the odds of a positive test result in a patient with disease compared with a patient without disease). A curve can be plotted on the ROC plot that corresponds with values of sensitivity and specificity which all have the same DOR. The resulting SROC curve represents the overall test performance or DOR, allowing for variation in threshold.

We used the method proposed by Moses and colleagues[187] to fit both symmetric and asymmetric SROC curves. Asymmetric curves allow for an increasing or decreasing trend in DORs with a proxy measure of threshold[187] (see Appendix 5 for full details). The method considers the relationship between the logarithm DOR (denoted by D) and a summary measure of diagnostic threshold (denoted by S). The parameter S represents that variation in the DOR due to different thresholds. As a diagnostic threshold decreases, the numbers of positive diagnoses (both correct and incorrect) increases, and the measure of threshold increases.

In the equations and figures which follow, the logarithm of the DOR is denoted by D and the logarithm of the measure of threshold by S. D and S can be calculated using any of the equivalent equations:

$$S = \ln\left[\frac{TPR}{(1-TPR)} \times \frac{FPR}{(1-FPR)}\right] = \text{logit}(TPR) + \text{logit}(FPR)$$

$$D = \ln(\text{DOR}) = \ln\left[\frac{TPR}{(1-TPR)} \times \frac{(1-FPR)}{FPR}\right] = \ln\left(\frac{\text{LR +ve}}{\text{LR –ve}}\right) = \text{logit}(TPR) - \text{logit}(FPR)$$

where the logit indicates the logarithm of the odds, as used in logistic regression.

Moses and colleagues' method first considers a plot of D against S calculated for each of the studies and then computes the best-fitting straight line through the points on the graph. The equation of the fitted line is

$$D = a + bS$$

Testing the significance of the estimate of the slope parameter b indicates whether there is significant trend in diagnostic performance with threshold. The parameters D and S can then be combined to produce an SROC curve (Appendix 5). The value of a and its associated 95% CIs can be exponentiated to give the summary DOR and CIs. For a symmetrical curve, this value applies across the curve. For an asymmetric curve it is the DOR where sensitivity = specificity ($S = 0$). In addition to determining the DOR, it is helpful to re-express values from the curve as sensitivities and specificities. We have chosen to identify the value on the curve at the mean value of S which lies somewhere close to the centre of the data. This value should be indicative of the average sensitivity and specificity, but does not account for the variability in values between studies.

Heterogeneity and subgroup analyses

Within each review, sources of heterogeneity were investigated by adding covariates to the standard Moses regression model. The exponential of the resulting coefficients for each of these terms gives the relative diagnostic odds ratio (RDOR) in each subgroup relative to the rest, or the ratio of the DOR in one subgroup of studies compared with the DOR in the other subgroup(s). Sources of heterogeneity investigated were as follows:

1. test used, such as AMTD, Amplicor, LCx, Amplicis Myco B and in-house for NAAT tests
2. study setting: laboratory versus hospital-based
3. reference standard used: culture plus clinical suspicion with or without additional tests versus other reference tests
4. study design: prospective versus retrospective/unknown
5. quality criteria:
   (a) patients representative versus unrepresentative/unknown
   (b) index blinded versus not blinded/unknown
   (c) reference blinded versus not blinded/unknown
6. prevalence of TB:
   (a) prevalence <10% versus prevalence 30+%
   (b) prevalence 10–20% versus prevalence 30+%
   (c) prevalence 20–30% versus prevalence 30+%
7. % patients smear positive:
   (a) <5% smear positive versus >30%
   (b) 5–10% smear positive versus >30%
   (c) 10–20% smear positive versus >30%
   (d) 20–30% smear positive versus >30%.

The overall quality of the specimens examined, which varies considerably, was not specifically assessed, principally, for example minimum volume CSF sample, as there was often little information provided from which to make an informed conclusion.

We used five study design-related criteria to examine whether accuracy estimates were altered in the subgroups of studies meeting two or more of these criteria:

1. studies meeting two design-related criteria (hospital-based and combined reference standard used)
2. studies meeting the two design-related criteria from 1 above, plus index and reference test blinded, and patient sample judged to be representative.

Chapter 4

Results of literature search and study screening

Available systematic reviews

No systematic reviews evaluating the accuracy of diagnostic tests for the investigation of TB infection were identified from the literature searches. Following completion of the searches, several systematic reviews have been identified:

- one of PCR for detection of smear-negative pulmonary TB infection[188]
- one of NAATs for detection of tuberculous pleuritis[59]
- one of NAATs for detection of tuberculous meningitis[56]
- two of ADAs for detection of TB pleurisy,[86,189] one of which also assessed interferon-γ[86]
- one of interferon-γ tests for detection of active and latent TB infection.[52]

A summary of the five systematic reviews in active TB infection is provided in Appendix 6. Pertinent aspects of the results and conclusions of these reviews are discussed in the relevant results chapters below.

Available primary studies

The titles and abstracts of 6006 papers were screened and 2458 potentially eligible studies were identified. The full texts of 1822 papers from the electronic searches plus 74 identified from other sources were retrieved for more detailed evaluation; 562 papers from the electronic sources were published in non-English languages and the decision was taken not to retrieve them. *Figure 1* provides a flowchart of the screening process and reasons for exclusion of 1661 papers.

A total of 215 papers reporting complete 2 × 2 table data from 212 studies were included; these provided 368 data sets, each comparing an index test against a reference standard at a given cut-off. A breakdown of these data sets according to type of TB investigated and test evaluated is provided in *Table 1*. Eight individual types of TB were investigated using nine different types of test. A further group of studies investigated the tests in miscellaneous extrapulmonary samples. The majority of the data sets pertain to nucleic acid amplification tests (56%) and more than half of those (53%) were evaluated for use in suspected pulmonary TB infection. The next most commonly investigated single forms of TB infection were pleural TB and TB meningitis, with ADA tests and NAAT tests, respectively, the most commonly evaluated within those TB types. Further details of the tests evaluated are provided in the following sections.

A further 20 papers from 19 studies assessing fully automated liquid culture were included in the review.

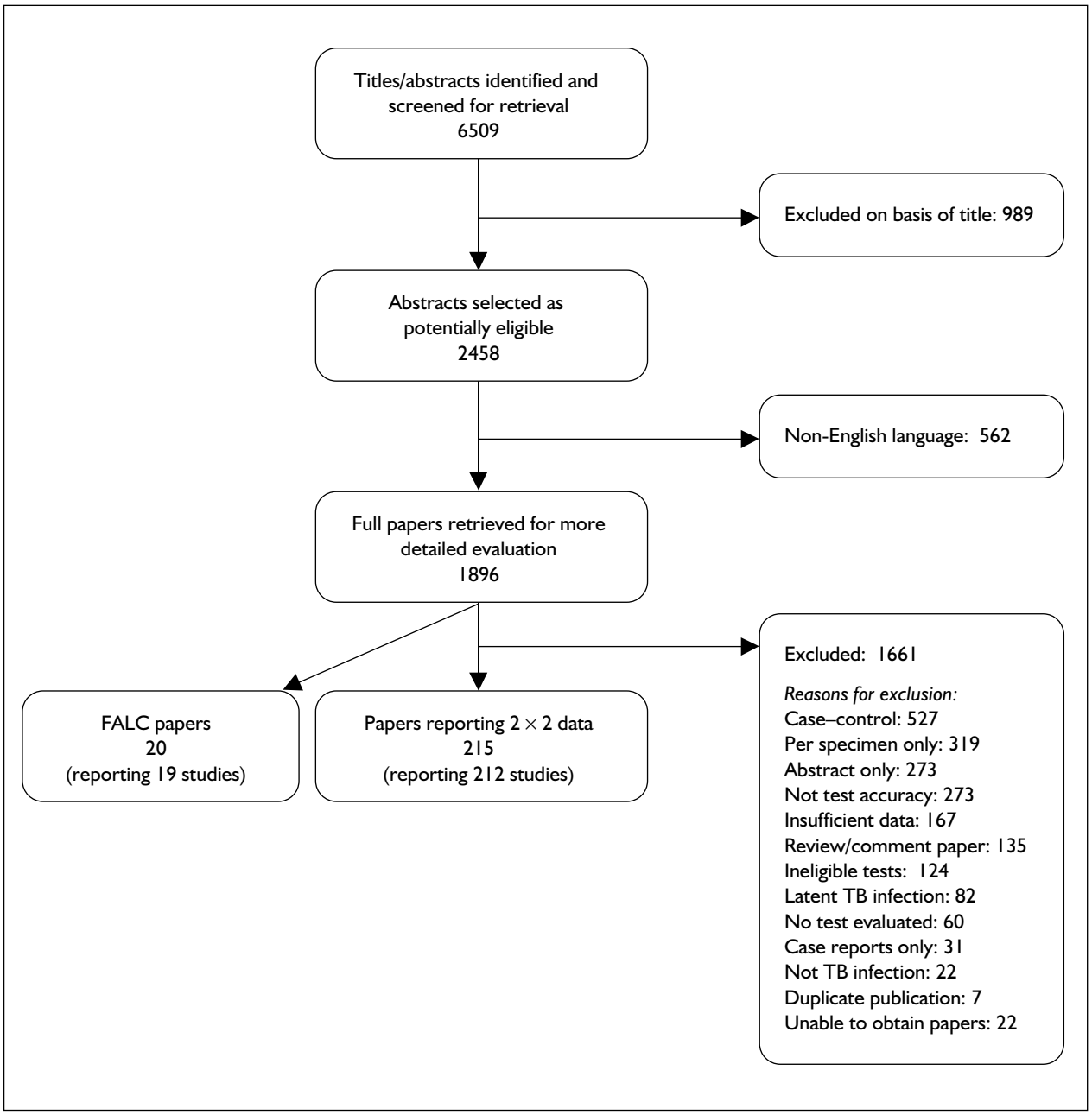


FIGURE 1 *Flowchart of screening process (ATB infection)*

TABLE 1 *Summary of all data sets included in primary analysis (one cut-off per study)*

	All data sets	Pulmonary TB	Pleural TB	TB meningitis	Lymphatic TB	Peritoneal TB	Pericardial TB	Genito-urinary TB	Skeletal TB	Miscellaneous extrapulmonary TB	Serum samples
Total	368	146	100	42	18	14	7	10	1	30	40
NAAT tests	207	110	20	26	16	1	2	9	1	22	3
Commercial	106	59	6	13	9	0	1	5	0	13	1
In-house	101	51	14	13	7	1	1	4	1	9	2
Molecular probe tests	5	4	0	0	0	0	0	0	0	1	0
Phage tests	5	5	0	0	0	0	0	0	0	0	0
Anti-TB antibody tests	39	21	7	4	1	0	0	0	0	6	25
Commercial	20	10	6	0	0	0	0	0	0	4	14
In-house	19	11	1	4	1	0	0	0	0	2	12
Adenosine deaminase tests	65	2	42	8	0	8	4	0	0	1	6
Antigen tests	6	1	1	2	1	0	0	1	0	0	1
Cytokine tests	18	2	13	0	0	2	1	0	0	0	3
Lysozyme tests	11	0	11	0	0	0	0	0	0	0	1
Other miscellaneous tests	12	1	6	2	0	3	0	0	0	0	1

Chapter 5

Results: detection of pulmonary tuberculosis disease

We identified a total of 147 test comparisons for pulmonary TB; 121 tests applied in respiratory specimens, 22 in serum or peripheral blood samples, two in pleural fluid and one using urine specimens (see *Table 2*) (the two data sets using pleural fluid specimens have been included in the pulmonary TB section as the study authors[190] specifically state that "patients suspected of pulmonary TB" were enrolled and furthermore they include "pleural effusion" as a

TABLE 2 *Pulmonary TB – summary of data sets identified*

	All	Respiratory	Serum	Other
Total	146	120	23	3
NAAT tests	**110**	**106**[a]	**4**	**0**
Commercial	59	59	0	0
AMTD	23	23	0	0
Amplicor	30	30	0	0
LCx	5	5	0	0
Amplicis (Myco B)	1	1	0	0
In-house	51	47	4	0
IS6110	30	28	2	0
Other targets	21	19	2	0
Molecular probe tests	**4**	**4**	**0**	**0**
BD ProbeTec	3	3	0	0
In-house	1	1	0	0
Phage tests	**5**	**5**	**0**	**0**
FastPlaqueTB	4	4	0	0
PhageTek MB	1	1	0	0
Anti-TB antibody tests	**21**	**3**	**17**	**1** urine
Commercial	10	2	8	0
Anda TB IgG	2	1	1	0
Anda TB IgM	2	1	1	0
Detect TB	1	0	1	0
EIA Pathozyme TB complex	1	0	1	0
ICT	2	0	2	0
Mycodot	2	0	2	0
In-house	11	1	9	1
14 kDa	1	0	1	0
38 kDa	1	0	1	0
ESAT-6	1	0	1	0
H37Ra	1	0	1	0
LAM	3	0	2	1
PPD	1	0	1	0
Sonicated MTB (unspecified)	3	1	2	0
Adenosine deaminase tests	**2**	**0**	**2**	**0**
ADA	1	0	1	0
ADA2	1	0	1	0
Antigen tests	**1**	**1**	**0**	**0**
LAM & H37Rv	1	1	0	0
Cytokine tests	**2**	**0**	**0**	**2** pleural fluid
IFN gamma	1	0	0	1
Interleukin	1	0	0	1
Other miscellaneous tests	**1**	**1**	**0**	**0**
TBSA	1	1	0	0

[a] Includes two data sets in gastric aspirate only (all Amplicor test).

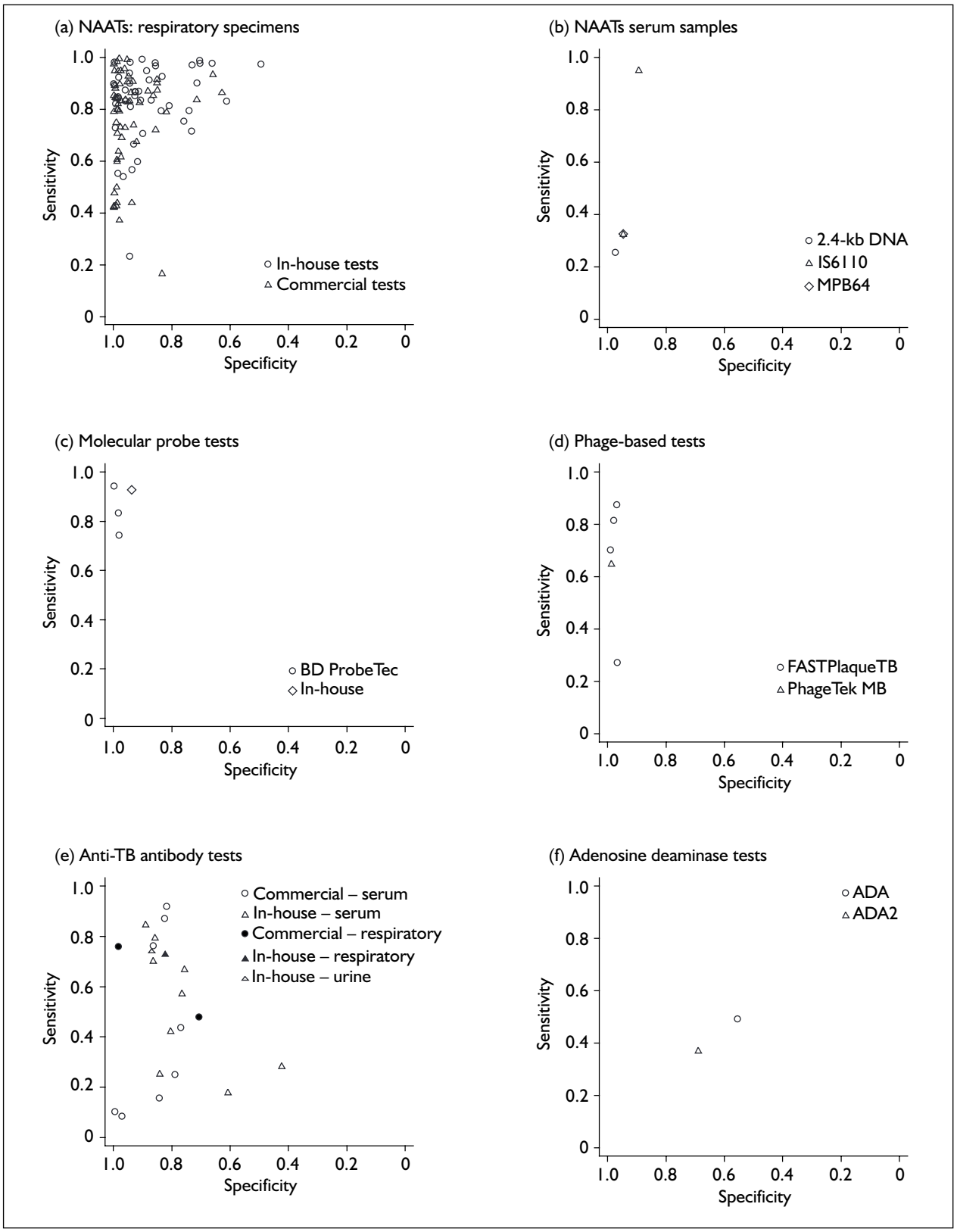


FIGURE 2 *Pulmonary TB – ROC plots for each group of tests*

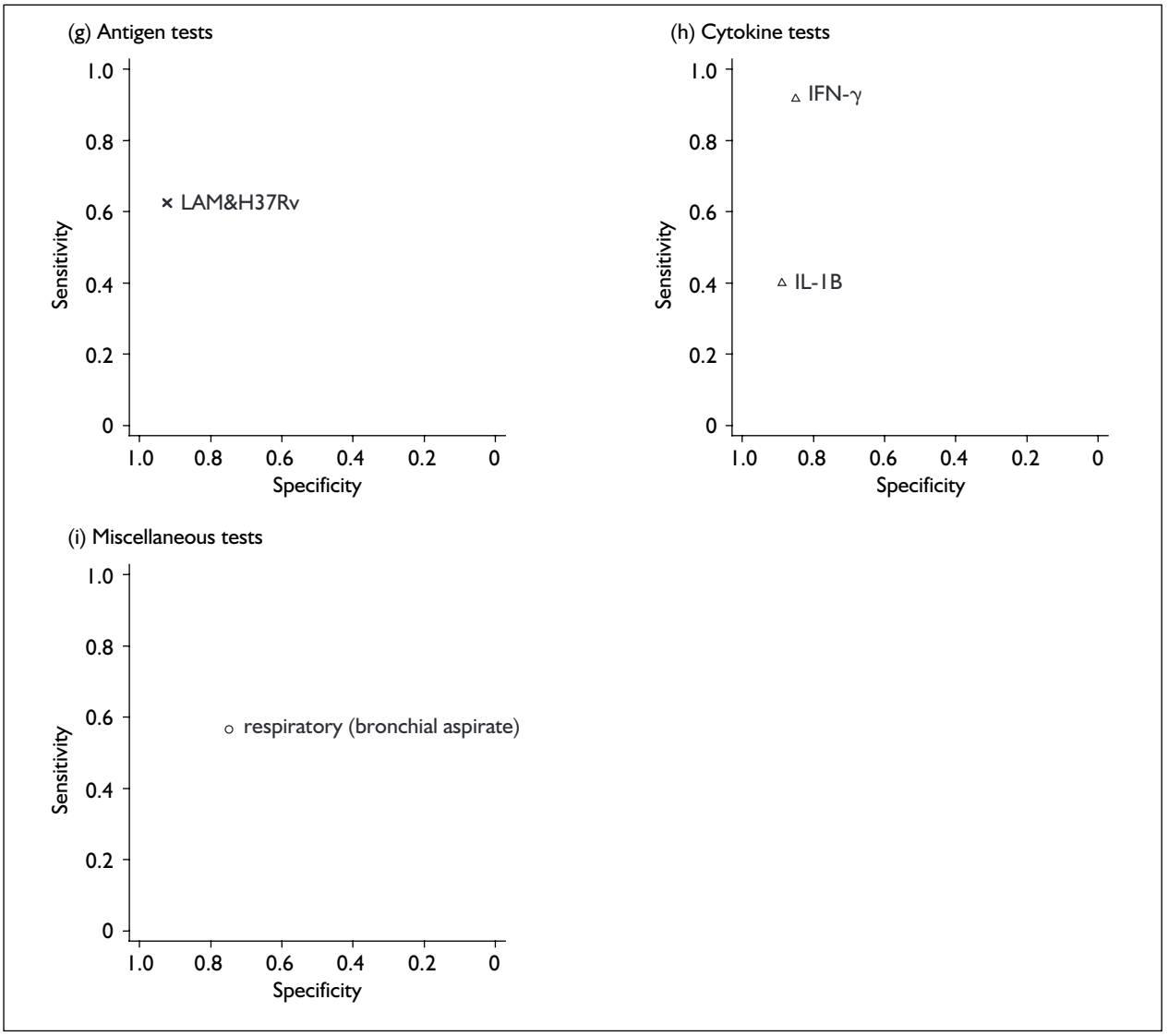


FIGURE 2 *Pulmonary TB – ROC plots for each group of tests (continued)*

respiratory specimen). ROC plots of each sensitivity and specificity pair for each group of tests are provided in *Figure 2(a–i)*.

Nucleic acid amplification tests

Description and quality of included studies

We included 110 data sets relating to NAATs; 106 applied to respiratory specimens (two used on gastric aspirates only) and four applied to peripheral blood samples. A summary of key characteristics across all data sets is given in *Table 3*, with details per study provided in Appendix 7 grouped according to evaluation of commercial or inhouse tests.

Tests

More than half (59/110) of the evaluations related to commercial NAATs; the remainder were 'in-house' tests, developed and used in individual laboratories. The commercial test evaluations were predominantly of the Gen-Probe Amplified Mycobacterium Direct Test® (AMTD) (n = 23) or the Roche Amplicor® MTB test (n = 30), with five data sets relating to the Abbott Laboratories LCx test, which has recently been discontinued.

Of the 51 in-house test evaluations, 30 were based on the IS6110 target sequence. The next most commonly used target was the gene encoding the 65-kDa antigen (five studies) and then MTP40 or IS986 (two studies each).

TABLE 3 *Diagnostic test evaluations in pulmonary TB – summary of key study characteristics for main tests*

	NAAT	**Amplification plus probe**	**Phage**	**Anti-TB antibody**
Total no. of studies	106	4	5	21
Mean sample size	269 (SD 399); range 14–3794	213 (SD181); range 30–402	606 (SD 564); range 63–1483	304 (SD 284); range 54–1000
Commercial	59	3	4	10
Mean sample size	362 (SD 506); range 22–3794			250 (SD 206); range 73–593
In-house	47	1	1	11
Mean sample size	153 (SD 128); range 14–833			353 (SD344); range 54–1000
Reference standard				
Culture + clinical diagnosis ± other	46 (43%)	1 (25%)	2 (40%)	8 (38%)
Culture + anti-TB therapy	2 (2%)	1 (25%)	0	0
Culture + other	0	0	0	5 (24%)
Culture alone	52 (49%)	2 (50%)	3 (60%)	5 (24%)
Clinical diagnosis alone	3 (3%)	0	0	3 (14%)
Disease prevalence (mean, SD, range)	29.2% (SD 19.3); range 2–78%	29.7% (SD 19.1); range 7–47%	32% (SD 18); range 14–52%	24% (SD 12); range 9–53%
Setting				
Hospital-based	77 (73%)	0	3 (60%)	20 (95%)
Clinic	0	0	1 (20%)	0
Laboratory-based	26 (25%)	3 (75%)	1 (20%)	0
Unknown	3 (3%)	1 (25%)	0	1 (5%)
Sample type				
>95% sputum samples	40 (38%)	0	5 (100%)	3 (14%)
No sputum samples included	10 (9%)	1 not reported	0	0
Serum	0	0	0	17 (81%)
Urine	0	0	0	1 (5%)
Patients representative?				
Yes	62 (58%)	2 (50%)	3 (60%)	14 (67%)
No	14 (13%)	1 (25%)	1 (20%)	2 (10%)
Unclear	30 (28%)	1 (25%)	1 (20%)	5 (24%)
Study design prospective?				
Yes	26 (25%)	0	0	14 (67%)
No	8 (8%)	1 (25%)	1 (20%)	0
Unclear	72 (68%)	3 (75%)	4 (80%)	7 (33%)
Index test interpreted blinded?				
Yes	29 (27%)	0	1 (20%)	8 (38%)
No	1 (1%)	0	0	0
Unclear	76 (72%)	4 (100%)	4 (80%)	13 (62%)
Reference test interpreted blinded?				
Yes	27 (25%)	0	1 (20%)	6 (29%)
No	2 (2%)	0	0	0
Unclear	77 (73%)	4 (100%)	4 (80%)	15 (71%)

Reference tests

Of the 106 data sets on respiratory specimens, 52% (55/106) used culture alone as the reference standard, 44% ($n = 46$) combined culture with clinical symptoms (with or without an assessment of response to anti-TB therapy or chest X-ray), two used culture plus treatment response and three used clinical diagnosis plus treatment response only.

Sample details

Most of the studies in this section (79/106) were hospital based, that is, recruited referred patients. A further 26 were 'laboratory based' (recruiting samples rather than patients) and in three cases the source of the population was not clear. In 40 studies (38%), more than 95% of specimens analysed were sputum, and in 10 (9%), no sputum samples were included [samples were gastric

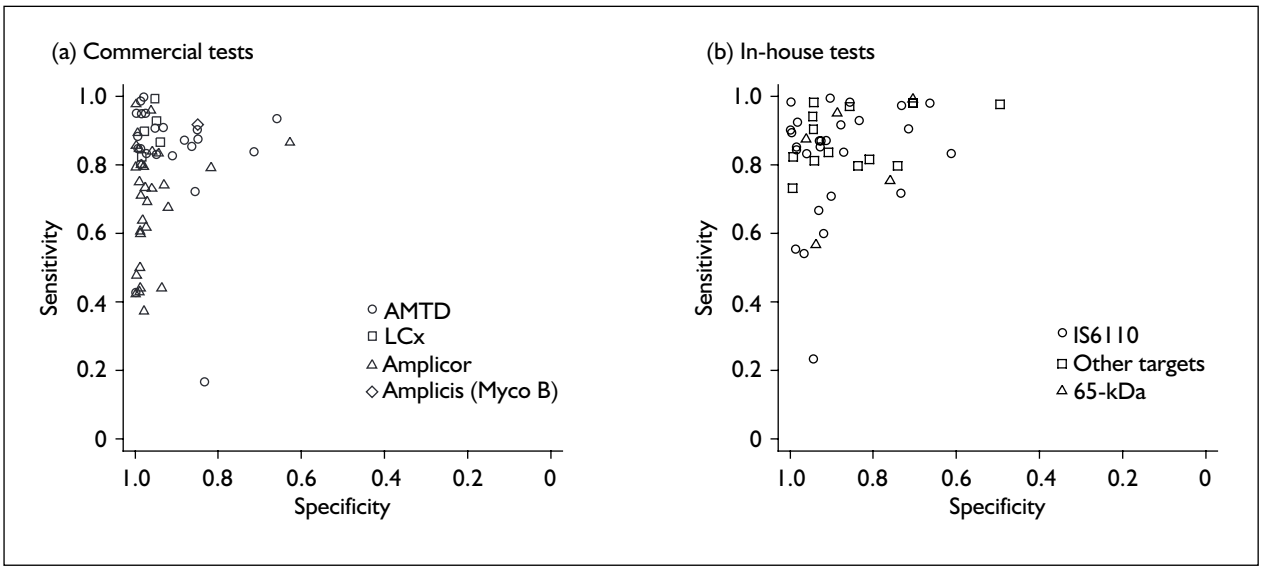


FIGURE 3 *NAAT evaluations in pulmonary TB (respiratory specimens) – ROC plots by test*

aspirate (2), BAL (5), lung aspirate (from percutaneous transthoracic needle aspiration) (1)] and in the remaining studies mixed specimens were included.

Across all 106 data sets, the mean number of patients recruited was 269 [standard deviation (SD) 399; range 14–3794]. The mean was much lower for the studies of inhouse tests (153, SD 128, range 14–833) compared to commercial tests (n = 362, SD 506; range 22–3794). Mean prevalence of TB was 29% (SD 19%).

In 58% (62/106) of studies, we judged the patient sample to have been representative (i.e. the study at least stated that patients 'suspected' of having TB were recruited). The sample was judged to be unrepresentative in 14 studies and was unclear in 30 studies.

Test interpretation

Index test interpretation was reported as blinded in 29/106 studies and not blinded in one study and was unclear in 76 studies. Twenty-six of the studies were clearly prospective in design and eight were retrospective. Assuming (based on time taken to perform PCR compared with culture) that the index test was interpreted first in the prospective studies and was therefore to all intents and purposes 'blinded', increases the number of studies in which the index test was interpreted blinded to 44. The reference standard was clearly reported as interpreted blinded in 27 studies, 20 of which also reported blinded index test interpretation.

Results

Sensitivity and specificity

The plot of sensitivity against specificity for each study according to main specimen type is presented in *Figure 3(a)* and *(b)*. For the respiratory specimen studies, there is a considerable range in both sensitivity and specificity estimates, for both commercial and in-house tests. This is confirmed statistically by the heterogeneity tests reported in *Table 4*, which were all statistically significant ($p < 0.01$) except for both sensitivity and specificity of the LCx test. The ROC plot of the commercial group by test (*Figure 3a*) shows that the five LCx data sets are in fact clustered together in the top left-hand corner of the plot, so the statistical significance of the result is unlikely to be due to the low power of the test in such a small sample of studies.

This plot also shows that the specificity of the Amplicor test in most studies is over 90% but the sensitivity varies to a much greater extent, from just under 40 to 100%.

Heterogeneity investigation (respiratory specimens only)

Before pooling all data sets together, we investigated whether selected variables had any impact on overall test accuracy by adding covariates for each variable of interest to a regression model (*Table 5*). The results demonstrate that none of the commercial tests are significantly more or less accurate than the in-house tests and, for the in-house tests, the use of the target sequence IS6110 has no significant

TABLE 4 *NAAT evaluations in pulmonary TB – tests for heterogeneity of sensitivities and specificities*

Comparison	*n*	Sensitivity		Specificity	
		Q statistic	*p*-Value	*Q statistic*	*p*-Value
All NAAT data sets (respiratory specimens)	106	657	<0.01	1634	<0.01
Commercial tests	**59**	**378**	**<0.01**	**1158**	**<0.01**
AMTD	23	85	<0.01	454	<0.01
Amplicor	30	136	<0.01	203	<0.01
LCx	5	5	0.27	8	0.09
Amplicis (Myco B)	1	NA		NA	
In-house tests	**47**	**245**	**<0.01**	**363**	**<0.01**
IS6110	28	172	<0.01	177	<0.01
Other targets	19	65	<0.01	153	<0.01
All NAAT data sets (serum samples)	5	28	<0.01	3	0.35

NA, not applicable.

TABLE 5 *NAAT evaluations in pulmonary TB – regression analyses to identify source(s) of heterogeneity*

Comparison	*n* (106)	Model parameters (95% CI) and *p*-value		
		Coefficient	*p*-Value	RDOR[a]
Test type				
AMTD vs inhouse	23/47	0.43	0.38	1.54 (0.60 to 3.98)
Amplicor vs inhouse	30/47	0.00	0.99	1.00 (0.37 to 2.69)
LcX vs inhouse	5/47	1.24	0.16	3.45 (0.61 to 19.56)
Amplicis Myco B vs inhouse	1/47	–0.35	0.85	0.70 (0.02 to 29.33)
For in-house only: IS6110 vs other targets	28/19	0.16	0.77	1.17 (0.39 to 3.54)
Institute				
Laboratory-based vs hospital-based	26/80	1.84	<0.01	6.28 (2.95 to 13.34)
Reference standard				
Culture plus clinical vs culture alone	46/55	–0.83	0.03	0.44 (0.21 to 0.90)
Culture plus other vs culture alone	2/55	–0.93	0.48	0.40 (0.03 to 5.41)
No culture vs culture alone	3/55	–1.36	0.21	0.26 (0.03 to 2.20)
Design				
Prospective vs retrospective/unknown	26/80	–0.31	0.46	0.73 (0.32 to 1.69)
Quality factors				
Patients representative vs unrepresentative/ unknown	62/44	–0.12	0.74	0.88 (0.43 to 1.82)
Index test blinded vs not blinded/unknown	29/77	–1.09	0.01	0.34 (0.15 to 0.73)
Reference blinded vs not blinded/unknown	27/79	–1.13	0.01	0.32 (0.15 to 0.71)
TB prevalence				
Prevalence <10% vs prevalence 30+%	19/49	0.82	0.13	2.26 (0.79 to 6.49)
Prevalence 10–20% vs prevalence 30+%	21/49	0.68	0.19	1.96 (0.71 to 5.45)
Prevalence 20–30% vs prevalence 30+%	17/49	0.09	0.86	1.10 (0.39 to 3.12)
% patients smear positive				
<5% smear positive vs >30%	22/17	–0.98	0.10	0.37 (0.12 to 4.74)
5–10% smear positive vs >30%	18/17	0.31	0.62	1.36 (0.40 to 4.70)
10–20% smear positive vs >30%	19/17	–0.88	0.14	0.42 (0.13 to 1.34)
20–30% smear positive vs >30%	24/17	–0.60	0.28	0.55 (0.18 to 1.65)

[a] Relative diagnostic odds ratio, i.e. in one group compared with the other.

impact compared with the other target sequences used. A further regression analysis (not shown) indicates that the accuracies of the commercial tests were not significantly different from each other.

Laboratory-based studies and studies using culture alone as a reference standard were found to have significantly higher accuracy than hospital-based studies ($p < 0.01$) and studies using combined reference standards ($p = 0.03$). The higher

accuracy in the laboratory-based studies could be partially due to the use of culture alone as the reference standard in 17 of the 26 studies. It seems likely that the make-up of the populations included in the laboratory-based studies could additionally contribute to the difference in accuracy.

Studies explicitly using blinded index test interpretation and blinded reference standard interpretation had significantly lower accuracy than those that did not report using blinding ($p < 0.01$) (*Table 5*). Twenty studies blinded both index and reference test interpretation.

None of the other variables investigated (study design, patient representativeness, disease prevalence or prevalence of AFB smear-positive patients) statistically affected accuracy.

SROC analysis

Respiratory specimens

The overall pooled analysis indicates that in respiratory specimens, the DOR is 116.58 (95% CI: 77.04 to 176.42) with an associated sensitivity of 85.8% and specificity of 95.9% (*Table 6*). Amongst the commercial tests, the highest accuracy was seen with the LCx test (DOR 544.50, 95% CI: 109.69 to 2702.80), although the number of studies was small and CIs wide. The Amplicor test had much lower accuracy than both the AMTD test and the inhouse tests, primarily owing to lower sensitivity (73.0% compared with 88.3 and 89.2%, respectively). *Figure 4* plots a separate SROC curve per test.

Accuracy was lower in each subgroup of studies meeting each one of five design-related criteria. In particular, studies reporting blinded reference test interpretation and studies that were hospital based as opposed to laboratory based demonstrated much lower accuracy: DOR 51.95 (95% CI: 31.07 to 86.86) and 77.05 (95% CI: 53.85 to 112.33).

When more than one of the design-related criteria was applied together, overall accuracy progressively dropped. The DOR in studies meeting two criteria (hospital based and combined reference standard used) was 67.58 (95% CI: 38.34 to 119.10), and in those meeting all five criteria was 40.72 (95% CI: 15.23 to 108.84). Associated sensitivity estimates in these two groups were both around 79%, but specificity dropped from 95.3 to 92.8%.

Respiratory specimens by smear status

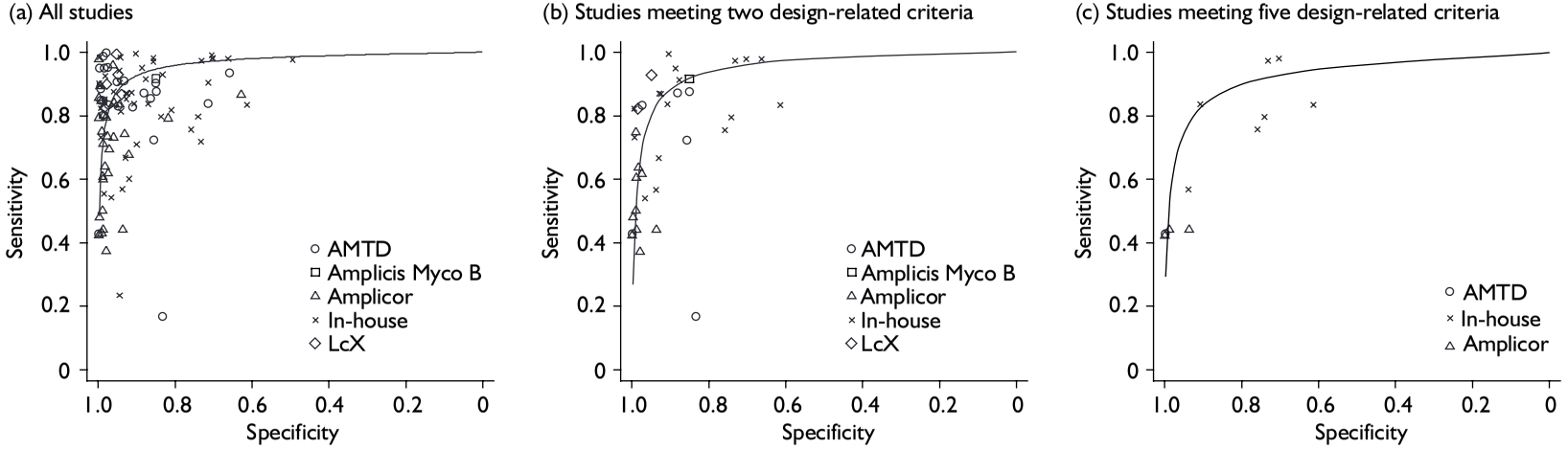
Around one-third of the studies provided separate accuracy data according to the smear status of the participants; 28 provided data for smear-negative patients and 25 for smear-positive group (in three studies, all recruited patients were smear negative). The individual study data in Appendix 7 show that several of the studies reporting smear-positive subgroup data actually included very few smear positive patients (mean 67, SD 101) and in many cases this resulted in either no or very few patients in the reference test negative group (mean 17, SD 65). As a result, a large proportion of studies reported either 100% sensitivity ($n = 8$) and/or 100% specificity ($n = 7$) or 0% specificity ($n = 6$). Small numbers of patients mean that when the zero cell correction (described in the section Data synthesis. p. 17) is applied, the resulting accuracy estimates are considerably affected. We therefore excluded all data sets with less than 10 diseased or non-diseased patients, leaving three data sets of smear-positive patients (mean sample size 251, SD 238) (*Figure 5a*) and 18 from smear-negative patients (mean sample size 192, SD 188) (*Figure 5b*).

The DOR from the SROC analysis (*Table 7*) for the smear-positive patients was much higher than for the smear-negative group [828.09 (95% CI: 0.20 to 3,464,261.37) compared with 35.80 (95% CI: 17.68 to 72.51)]. As might be expected, this was due to significantly lower sensitivity in the smear-negative subgroup (73.2% compared with 96.8% for the smear-positive group), confirming that NAATs are more likely to detect TB cases accurately when AFB are present in the sample.

Of the individual tests, for smear-positive patients the AMTD test performed extremely well in the single study in our sample,[191] with 100% sensitivity and 98% specificity (before zero cell correction). The Amplicor test was also shown to have high specificity (95.3%), but sensitivity was slightly lower at 89.5% (*Table 7*). Only one[192] of the three studies met our two main design-related criteria; the other Amplicor study[193] used culture alone as the reference standard and the AMTD study[191] was laboratory based. Sensitivity and specificity were 94 and 100%, respectively (Appendix 7).

For the smear-negative patients, of the commercial tests the AMTD test had the highest DOR and best combination of sensitivity (70.2%) and specificity (94.6%). The Amplicor test had similarly high specificity (95.5%) but much lower sensitivity (57.5%). Sensitivity in the five studies using in-house tests was much higher than for the commercial tests (91.0%); however, the mean sample size was only around half that of the commercial studies (92 compared with 230) and

TABLE 6 *NAAT evaluations in pulmonary TB (respiratory specimens) – SROC regression models and associated SROC curves*



	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se^d (%)	Sp^d (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se^d (%)	Sp^d (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se^d (%)	Sp^d (%)
All tests	106	116.58 (77.04 to 176.42)	0.06	85.8	95.9	37	67.58 (38.34 to 119.10)	0.41	79.9	95.0	12	40.72 (15.23 to 108.84)	0.19	79.4	92.8
AMTD	23	162.28 (50.42 to 522.33)	0.66	88.3	96.1	6	22.39 (0.68 to 733.62)	0.57	68.0	94.7	1	501.69 (29.54 to 8520.63)	NA	42.7	99.9
Amplicor	30	49.07 (13.73 to 175.38)	0.03	73.0	98.4	10	6.76 (0.53 to 86.08)	0.04	52.9	98.9	3	77.82 (0.54 to 11303.49)	NA	43.4	99.0
LcX	5	544.50 (109.69 to 2702.80)	0.19	93.3	96.5	2	265.20 (76.56 to 918.64)	NA	88.6	97.1	0				
Amplicis Myco B	1	62.70 (23.97 to 163.98)	NA	91.7	85.1	1	62.70 (23.97 to 163.98)	NA	91.7	85.1	0				
In-house	47	100.91 (59.27 to 171.81)	0.31	89.2	92.8	18	72.47 (34.84 to 150.75)	0.78	89.6	89.4	8	23.35 (9.00 to 60.62)	0.11	89.7	78.6
Subgroups meeting each design-related criterion															
Hospital-based	80	77.05 (52.85 to 112.33)	0.09	83.6	94.6										
Combined reference test	46	93.06 (51.64 to 167.70)	0.76	82.0	95.5										
Index blind/prospective	44	86.89 (49.55 to 152.36)	0.85	84.0	94.2										
Reference test blinded	27	51.95 (31.07 to 86.86)	0.16	81.3	93.2										
Patients representative	62	103.34 (62.16 to 171.79)	0.17	81.3	93.4										

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.

[b] DOR where sensitivity = specificity.

[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.

[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

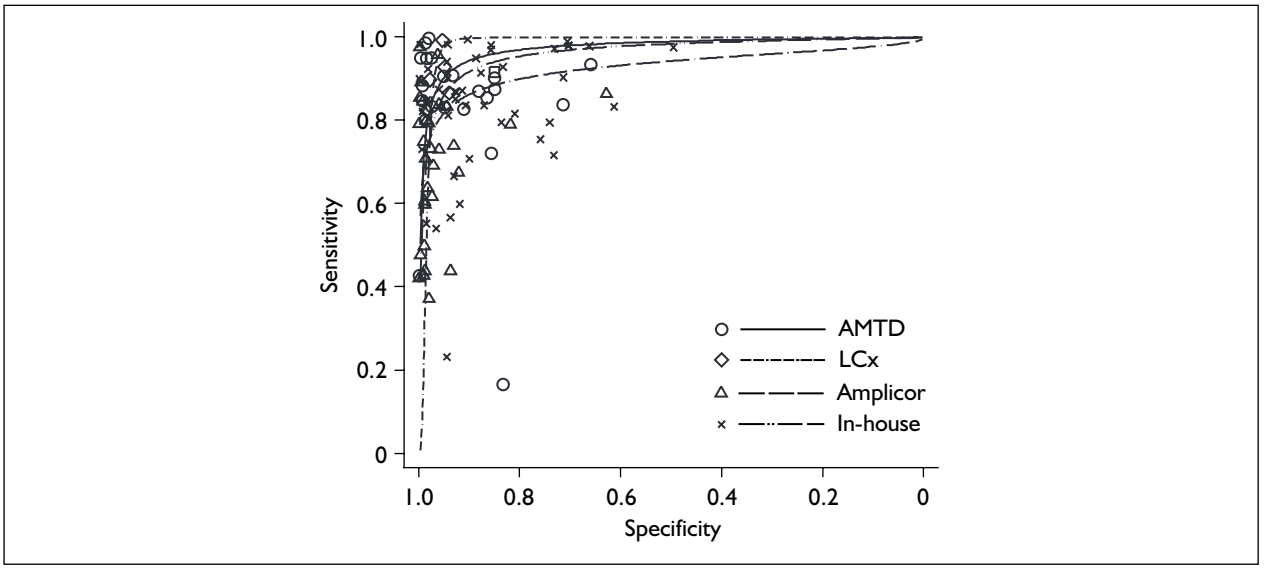


FIGURE 4 *NAAT evaluations in pulmonary TB (respiratory specimens) – SROC curves by test*

all of the studies used culture alone as the reference standard which may have contributed to higher sensitivity estimates. However, of the commercial test evaluations, nine of the 13 also used culture alone as the reference standard. Only three of the smear-negative data sets (all evaluating the Amplicor test) met our two main design-related criteria; sensitivity and specificity were 44.0 and 98.8%, respectively (Appendix 7).

Serum specimens

NAATs do not appear to perform well when used on blood samples. The overall DOR across the four studies was 19.92 (95% CI: 2.09 to 189.61) with an associated sensitivity of 52.8% and specificity of 94.7% (*Table 8*). This result, however, is dominated by three data sets from the same study,[194] each using a different target sequence but all of which had sensitivities ranging between 26 and 33%. The other study in this group[175] had a sensitivity of 95% and a specificity of 89% using the IS6110 target sequence with serum samples, although the sample size was small (n = 88).

Summary

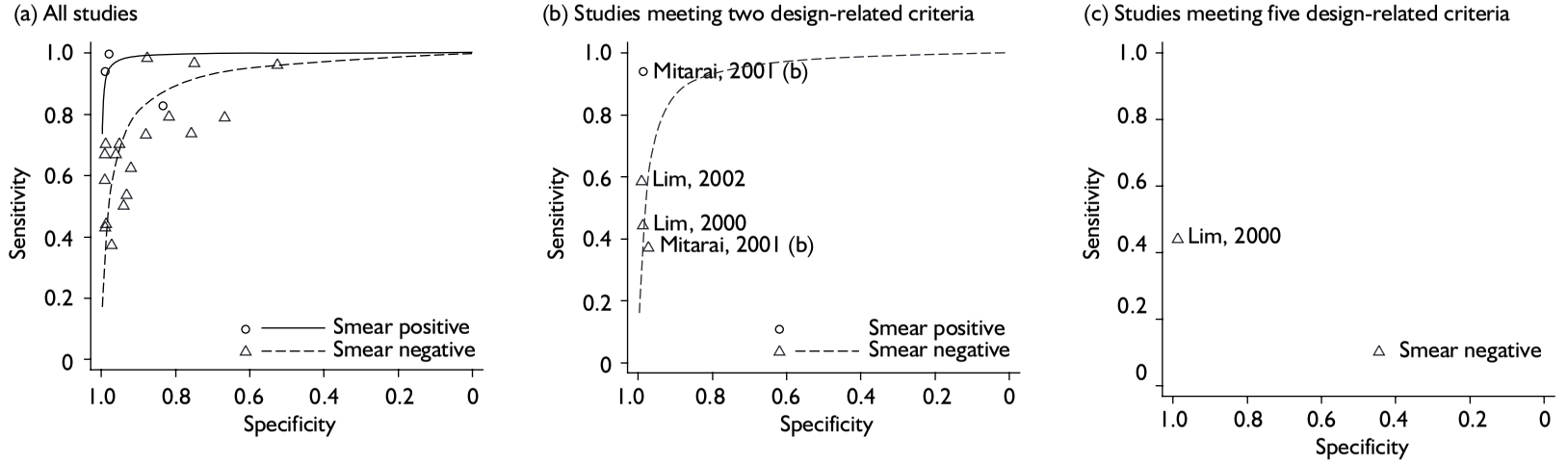
The sensitivity and specificity pairs for both commercial and in-house NAAT tests are generally all in the top left-hand quadrant of the ROC plot, but are clearly very heterogeneous. The type of test or target sequence used does not appear to explain this heterogeneity. We found the main explanatory factors for the variability to be the reference standard used, whether the study was laboratory or hospital based, and the use of blinded test interpretation, indicating that study design-related factors appear to have more impact on study results than patient- or setting-related factors.

Accuracy was significantly higher when culture alone was used as a reference standard. The sensitivity of culture is linked to the number of organisms present in the sample and so, as expected, positive culture rates will be lower in microscopy smear-negative sputum compared with smear-positive specimens; culture remains arguably the most sensitive laboratory detection method for active TB, although not all patients with clinical TB will have positive cultures.

Higher accuracy in the laboratory-based studies could be due partially to the greater use of culture alone as the reference standard in these studies and also to a bias resulting from the sample recruited to the study. Hospital-based studies will include specimens from all patients suspected of having TB infection, and therefore may include patients with a wider range of differential diagnoses, whereas laboratory-based studies tend to include all specimens submitted to a central laboratory over a given period – the reasons underlying referral of these specimens to a specialist laboratory may be less clear. It may also be that laboratory-based studies recruit patients for whom the index of suspicion for TB is much higher, or that specialist laboratories are simply better at performing the test.

Blinded interpretation of both index and reference tests was associated with lower accuracy, providing further weight to calls for improved study design and reporting. We would have

TABLE 7 *NAAT evaluations in pulmonary TB by smear status (respiratory specimens) – SROC regression models and associated SROC curves*



	n[a]	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	*n*[a]	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	*n*[a]	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
Smear positive	**3**	**828.09 (0.20 to 3,464,261.37)**	**NA**	**96.8**	**96.5**	**1**	**1274.29 (72.00 to 22,554.16)**	**NA**	**93.9**	**98.8**	**0**				
By test															
AMTD	1	18763.00 (1051.05 to 334,950.20)	NA	99.7	98.0	0					0				
Amplicor	2	173.97 (0.00 to >1,000,000)	NA	89.5	95.3	1	1274.29 (72.00 to 22,554.16)	NA	93.9	98.8	0				
Smear negative	**18**	**35.80 (17.68 to 72.51)**	**0.51**	**73.4**	**93.7**	**3**	**57.46 (4.76 to 693.99)**	**NA**	**46.4**	**98.5**	**1**	**64.59 (19.77 to 210.98)**	**NA**	**44.0**	**98.8**
By test															
AMTD	3	41.63 (0.96 to 1804.07)	NA	70.2	94.6	0					0				
Amplicor	9	10.98 (4.68 to 25.76)	0.01	59.0	95.5	3	57.46 (4.76 to 693.99)	NA	46.4	98.5	1	64.59 (19.77 to 210.98)	NA	44.0	98.8
LcX	1					0					0				
In-house	5	38.72 (6.16 to 243.48)	0.30	91.0	82.8	0					0				

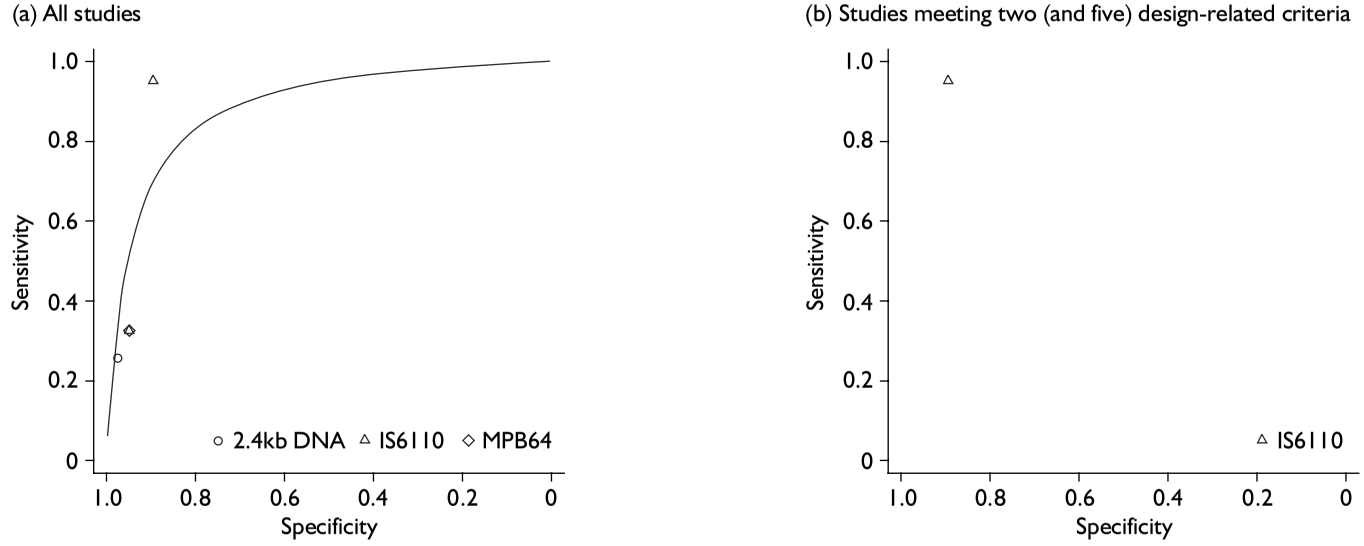
[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.

[b] DOR where sensitivity = specificity.

[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.

[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

TABLE 8 *NAAT evaluations in pulmonary TB (serum samples): SROC models and associated SROC curves*



Comparison	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
All	**4**	**19.92 (2.09 to 189.61)**	**NA**	**52.8**	**94.7**	**1**	**163.80 (30.02 to 893.76)**	**NA**	**95.1**	**89.4**
IS6110	2	37.73 (0.00 to >1,000,000)	NA	75.4	92.5	1	163.80 (30.02, 893.76)	NA	95.1	89.4
2.4-kb DNA	1	12.72 (2.88 to 56.19)		25.6	97.4	0				
MBP64	1	8.69 (2.88 to 26.19)		32.6	94.7	0				

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

expected to have seen lower accuracy in lower prevalence studies, but our results suggest that accuracy was actually higher in lower prevalence studies (although these results were not statistically significant). We can only assume that this was due to confounding, perhaps with better study design or less laboratory cross-contamination in lower prevalence studies.

Overall, test accuracy for NAAT tests is relatively good. Although the results were not statistically significant, the AMTD test appears to perform better than other currently available commercial tests. The LCx test actually performed the best, but it is no longer commercially available. When analyses were restricted to the better designed studies, overall pooled accuracy was still reasonably high, although sensitivity in particular is probably not high enough to permit accurate ruling out of TB infection (79.4%). The summary sensitivity for the commercial tests was particularly poor at only 42.7% for AMTD and 43.4% for Amplicor; specificities of both, however, were over 99%. The better quality in-house studies, interestingly, were much better at ruling out TB (sensitivity 89.7%), but less good at ruling it in (specificity 78.6%). However, these studies covered a greater range of different targets and procedures, so they were more heterogeneous compared with the commercial test evaluations.

The regression analyses did indicate that accuracy was, on the whole, lower in studies with lower proportions of smear-positive patients, as would be expected, although the results were not significant.

We were able to analyse separately three data sets restricted to smear-positive patients and 25 to smear-negative patients. The result for smear-positive patients indicate very high accuracy, as would be expected, that is, they can be used to rule in and rule out TB in most cases. The result for smear-negative patients demonstrated high specificity (93.7%) but lower sensitivity (73.4%), suggesting that NAAT tests can be used to rule in disease (i.e. a positive NAAT test result is unlikely to be a false negative), but owing to low sensitivity cannot be used to rule out disease.

Our results show that NAAT test accuracy progressively decreases in sputum smear-negative patients compared with sputum smear-positive, and in sputum culture-negative patients compared with culture-positive patients. Test accuracy and in particular test sensitivity are therefore related to the bacterial burden in clinical specimens. As a result, NAAT diagnostic sensitivity is insufficient to provide a reliable rapid rule out test in paucibacillary pulmonary TB, where the clinical need is greatest owing to the poor sensitivity of conventional diagnostic tests.

Simultaneous molecular amplification and probe tests

Description and quality of included studies

Five data sets evaluating molecular probe tests directly on clinical respiratory samples were included. A summary of key characteristics across all data sets is given in *Table 3*, with details per study provided in Appendix 8 grouped according to evaluation of commercial or in-house tests.

Tests

Three of the four data sets (75%) evaluated commercial test BDProbeTec ET, which simultaneously amplifies and detects samples in a closed homogeneous assay format (strand displacement assay).The other study[195] developed an in-house test combining PCR with a non-radioactive hybridisation technique. Each test was evaluated in a separate study.

Reference tests

Culture alone was used as the reference strategy in two of the studies (50%): one of the BDPRObeTec studies plus the in-house test evaluation. Only one study used the combination of culture with clinical TB diagnosis, and the remaining dataset used a culture plus anti-TB therapy trial.

Sample details

Three of the four studies were laboratory based; the study setting in the other was not clear. Three studies included a variety of respiratory samples and one included mixed respiratory and non-respiratory samples (Appendix 8), although more than half of included samples were sputum (52%).

The mean number of patients recruited to these studies was 213 (SD 181). The patients were judged to be representative in 50% of the studies (2/4).

Test interpretation

It was not possible to judge whether the index tests or reference standards had been interpreted blinded in any of the included studies. One study was retrospective in design and in the others the design was not clear.

TABLE 9 *Molecular probe evaluations in pulmonary TB – tests for heterogeneity of sensitivities and specificities*

		Sensitivity		Specificity	
Comparison	***n***	***Q* statistic**	***p*-Value**	***Q* statistic**	***p*-Value**
All molecular probe tests	4	9	0.03	5	0.20
BD ProbeTEc	3	8	0.02	3	0.24

Results

Sensitivity and specificity

The plot of sensitivity against specificity for each study is presented in *Figure 2(c)*. All points are clustered in the top left-hand corner of the plot, suggesting reasonable test accuracy. There is clearly more variation in sensitivity than specificity estimates and this is confirmed statistically by the heterogeneity tests reported in *Table 9*. The heterogeneity of the sensitivity estimates was statistically significant ($p = 0.03$) whereas the specificity estimates were homogeneous ($p = 0.20$). For the BDProbeTec test alone, sensitivity estimates were more heterogeneous ($p = 0.02$), but with only three studies the test has low power.

SROC analysis

The overall pooled analysis (*Table 10*) gives a DOR of 437.57 (95% CI: 31.66 to 6046.82) with an associated sensitivity of 88.1% and a specificity of 98.3%. The DOR for the BDProbeTec test was much higher at 572.86 (95% CI: 4.98 to 65941.94).

None of the studies in this group met our two chosen design-related criteria (hospital based and combined reference standard used). Owing to the small number of studies identified, we did not attempt to investigate further any potential causes of heterogeneity in this group.

Summary

There is insufficient evidence on which to make any clear recommendations on the use of molecular amplification with simultaneous probe tests directly in clinical specimens. Specificity is generally good, especially for BD ProbeTec ET, but sensitivity is more variable. None of the studies used blinded test interpretation and two of the studies used inappropriate reference standards, both of which are factors likely to inflate test accuracy considerably.

Phage tests

Description and quality of included studies

Five data sets evaluating phage-based tests reported in five separate studies were included. A summary of key characteristics across all data sets is given in *Table 3*, with details per study provided in Appendix 9.

Tests

Four of the five included studies (80%) evaluated the FastPlaque test; the fifth was of the PhageTek test.

Reference tests

Culture alone was used as the reference strategy in 60% ($n = 3$) of the studies (two FastPlaque and one PhageTek). The remaining two studies combined culture with clinical TB diagnosis, treatment response and X-ray.

Sample details

Only one of the studies was laboratory based, one was primary care clinic based and three were hospital based (*Table 3*). Four studies included only sputum samples and the fifth included a variety of respiratory samples, with 75% being sputum samples (Appendix 9).

The mean number of patients recruited to these studies was 606 (SD 564). About 49% (1483/3033) of the total patients were included in the study of PhageTekMB. The patients were judged to be representative in 60% of the studies (3/5).

Test interpretation

It was not possible to judge whether the index tests or reference standards had been interpreted blinded in any of the included studies. One study was retrospective in design and in the others the design was not clear.

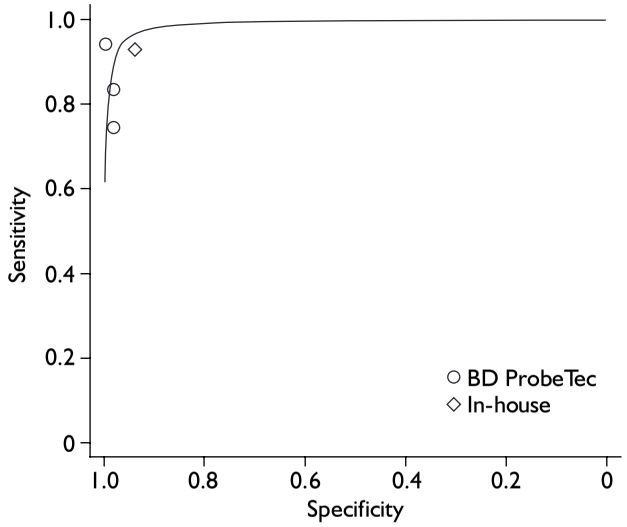
Results

Sensitivity and specificity

The plot of sensitivity against specificity for each study is presented in *Figure 2(d)*. Specificity was uniformly high across all five studies but sensitivity estimates were much more varied across the studies. The heterogeneity tests reported in *Table 11* indicate that the sensitivity estimates were statistically heterogeneous to $p < 0.01$, and the same was true for the FastPlaqueTB studies alone.

TABLE 10 *Molecular probe evaluations in pulmonary TB – SROC models and associated SROC curves*

(a) All studies



(b) Studies meeting two design-related criteria

None

Comparison	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
All tests[e]	**4**	**437.57 (31.66 to 6046.82)**	**NA**	**88.1**	**98.3**
BD ProbeTEc	3	572.86 (4.98 to 65941.94)	NA	86.0	98.9
In-house	1	195 (11.06 to 3438.0)	NA	92.8	93.7

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.
[e] Includes three pleural TB data sets, one peritoneal and two in pulmonary TB.

TABLE 11 *Phage-based test evaluations in pulmonary TB – tests for heterogeneity of sensitivities and specificities*

		Sensitivity		Specificity	
Comparison	***n***	***Q statistic***	***p*-Value**	***Q statistic***	***p*-Value**
All phage-based tests	5	46	<0.01	5	0.26
FastPlaqueTB	4	40	<0.01	4	0.26

SROC analysis

The overall pooled analysis (*Table 12*) gives a high DOR of 926.88 (95% CI: 5.34 to 106,998.45); however, despite a high associated specificity of 98.0%, the corresponding sensitivity was only 68.7%.

Only one study[196] of 781 patients met our first two quality criteria (hospital based and combined reference standard used); accuracy remained high.

Owing to the small number of studies identified, it was difficult to identify any potential causes of heterogeneity in sensitivity estimates. One potential factor could be the proportion of smear-positive specimens included in the studies. The two studies[196,197] that presented their results according to smear status (Appendix 9) found sensitivity to be much lower in smear-negative patients (49%[196] and 67%[197]) then smear-positive patients (84% and 87%, respectively). At the same time, specificity in smear-negative patients was extremely high in both studies (99%[196] and 98%[197]) and much lower in the smear-positive groups (89% and 88%, respectively), although patient numbers in the latter groups were very small.

Summary

The identified phage-based tests had very high specificity when applied directly in clinical samples, although sensitivity was lower and more variable. However, only five studies were available, limiting any conclusions that can be drawn.

Overall, there did not appear to be any clear pattern in sensitivity according to overall percentage of smear-positive patients included. However, this, along with the reference test used, could have been factors contributing to varying sensitivity. The main outlying study (sensitivity 27%) was also the smallest study,[198] with only 63 patients included. The authors suggest that duration of specimen storage and receipt of anti-TB therapy in the diseased group contributed to the low sensitivity. The two studies presenting results according to smear status suggest that the false negatives in the smear-positive groups were probably due to low AFB[197,199] and in some cases to the combined presence of *M. TB* and NTM.[197] The main reasons for the variation in sensitivity, however, will be biological, relating to the stability of the phage and the interaction of two living biological systems (the phage and the mycobacteria).

Anti-TB antibody tests

Description and quality of included studies

We included 21 data sets relating to anti-TB antibody serodiagnostic tests: three in respiratory specimens, one using urine specimens and 17 using serum samples. The 21 evaluations were carried out in 13 studies. A summary of key characteristics across all data sets is given in *Table 3*, with details per study provided in Appendix 10 grouped according to evaluation of commercial or in-house tests.

Tests

Just under half of the 21 evaluations (48%) were of commercial anti-TB antibody tests. The Anda TB IgG and IgM tests along with the ICT and Mycodot test were evaluated twice (*Table 2*). In the in-house test evaluations the most commonly evaluated (in three studies) antibody was lipoarabinomannan (LAM).

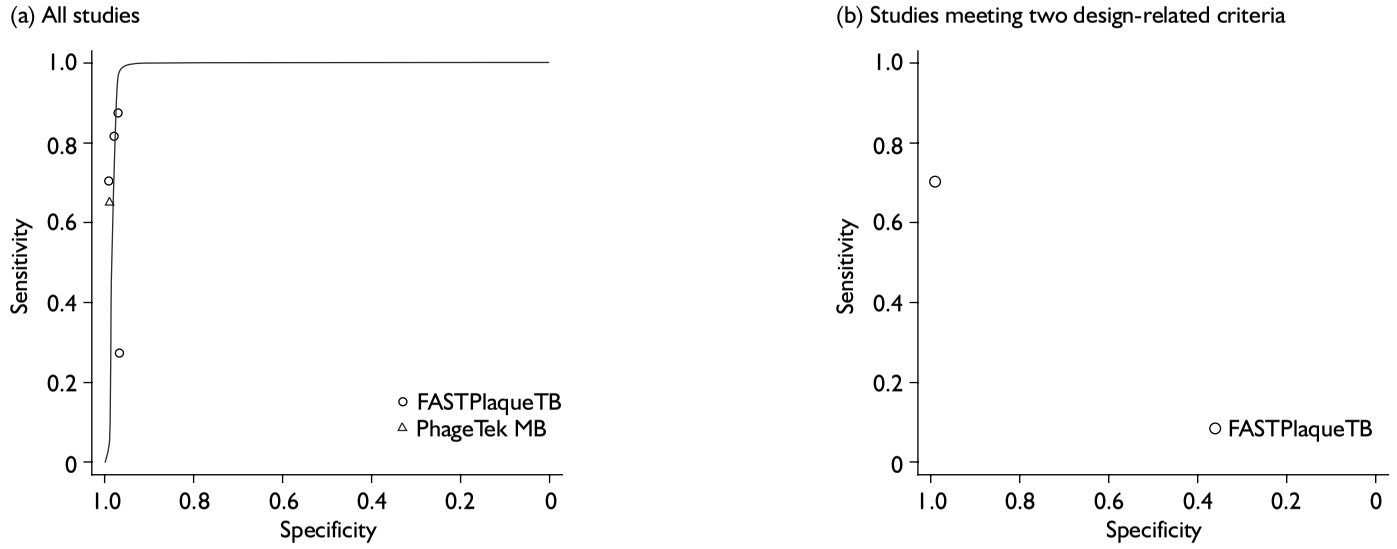
Reference tests

Culture alone was used as the reference strategy in five of the 21 evaluations (24%). Culture was combined with clinical diagnosis with or without additional tests in eight studies (38%) and clinical diagnosis alone used in three studies.

Sample details

About 95% of studies were reported to be hospital based and the patient sample was judged to be representative in 67% (*Table 3*). In most cases (17/21), serum samples were tested.

TABLE 12 *Phage-based test evaluations in pulmonary TB – SROC models and associated SROC curves*



Comparison	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
All tests[e]	**5**	**926.88 (5.34 to 160998.45)**		**0.26**	**68.7**	**98.0**				
FastPlaqueTB	4	100.72 (9.48 to 1070.14)	NA	69.6	97.8	1	223.87 (95.81 to 523.09)		70.3	99.0
PhageTek MB	1	145.04 (79.95 to 263.12)	NA	64.9	98.7					

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.
[e] Includes three pleural TB data sets, one peritoneal and two in pulmonary TB.

TABLE 13 *Anti-TB antibody test evaluations in pulmonary TB – tests for heterogeneity of sensitivities and specificities*

		Sensitivity		Specificity	
Comparison	***n***	***Q* statistic**	***p*-Value**	***Q* statistic**	***p*-Value**
All data sets[a]	21	330	<0.01	380	<0.01
By test					
Commercial	9	186	<0.01	65	<0.01
In-house	11	92	<0.01	282	<0.01
By sample					
Respiratory	3	5	0.10	10	0.01
Urine	1	NA	NA	NA	NA
Serum	17	284	<0.01	317	<0.01

[a] Includes three pleural TB data sets, one peritoneal and two in pulmonary TB.

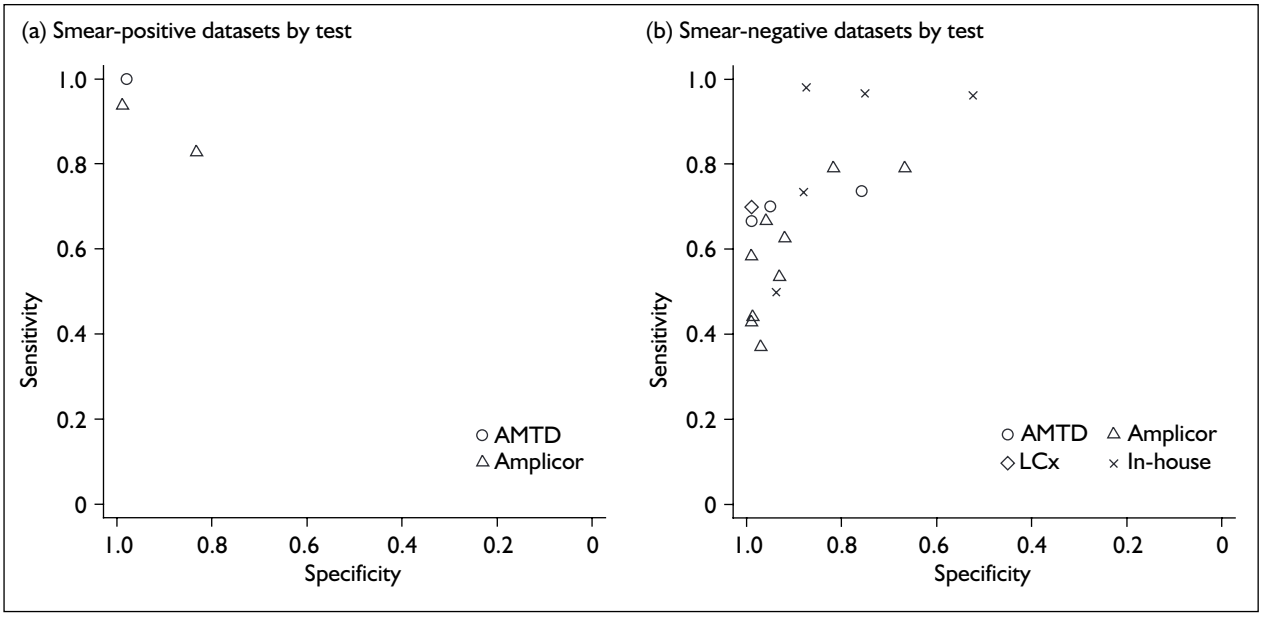


FIGURE 5 *NAAT evaluations in pulmonary TB (respiratory specimens) – ROC plots by smear status*

Across all data sets, the mean number of patients recruited was 304 (SD 284; range 54–1000). The mean was higher for the studies of in-house tests (353, SD 344) compared with commercial tests (250, SD 206). The mean prevalence of TB was 24% (SD 12%).

Test interpretation

Index test interpretation was reported as blinded in 38% (8/21) of studies and was unclear in remainder. Fourteen studies were clearly prospective in design. Assuming (based on time taken to perform serodiagnostic tests compared with culture) that the index test was interpreted first in the prospective studies and was therefore to all intents and purposes 'blinded', increases the number of studies in which the index test was interpreted blinded to 15 (71%). The reference standard was clearly reported as interpreted blinded in six studies, four of which also reported blinded index test interpretation.

Results

Sensitivity and specificity

The plot of sensitivity against specificity for each study according to main specimen type is presented in *Figure 2(e)*. The data sets are clearly very spread out across the plot. For all data sets together both sensitivity and specificity are statistically heterogeneous to $p < 0.01$ (*Table 13*) When separated according to specimen type and type of test, the studies in serum samples remain highly heterogeneous (*Table 13* and *Figure 6a* and *b*) although the sensitivities of data sets in respiratory

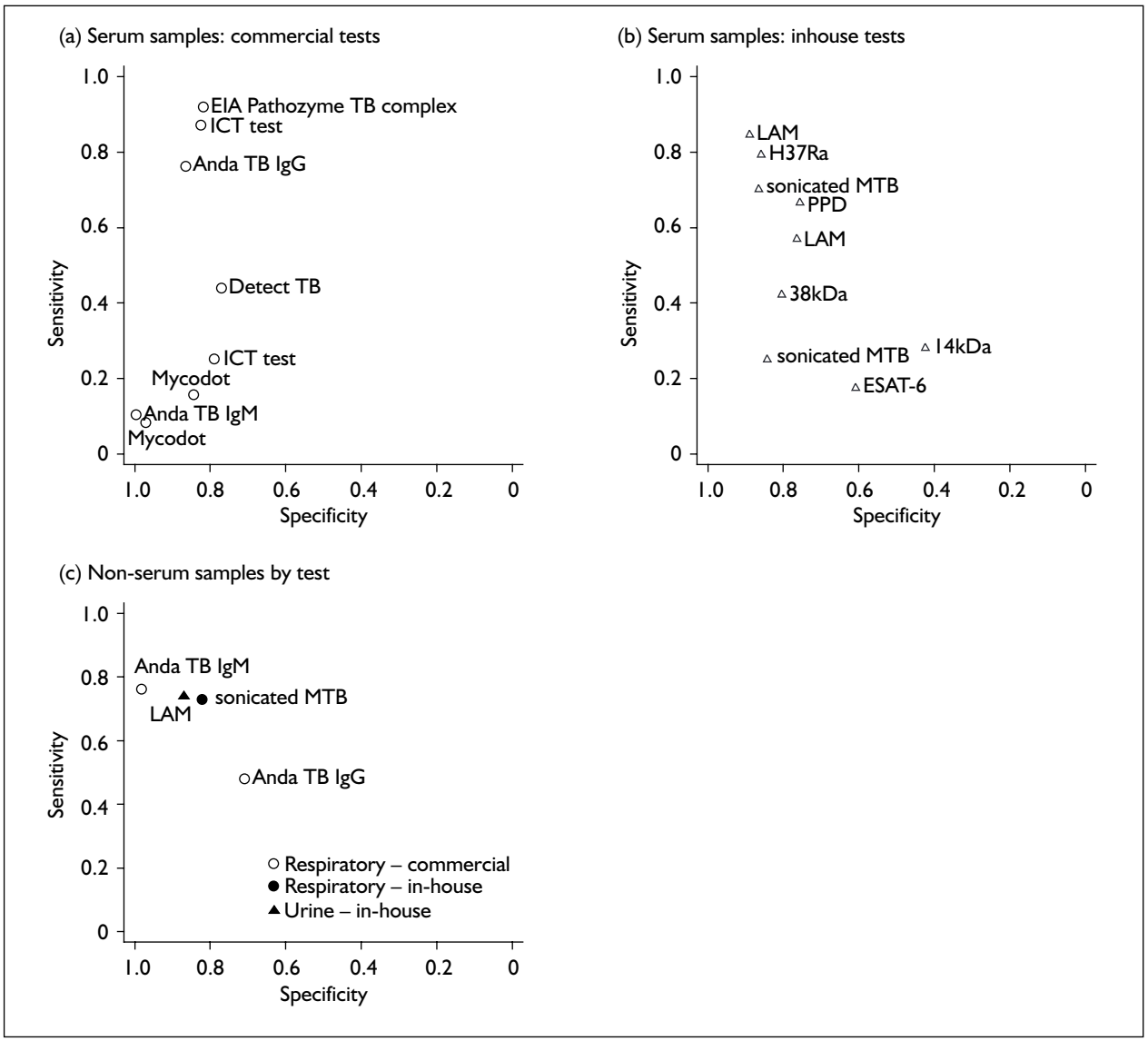


FIGURE 6 *Anti-TB antibody test evaluations in pulmonary TB – ROC plots*

samples are more homogeneous (*Figure 6c*). The plots do not appear to suggest the superiority of any one test or antibody over any of the others.

Heterogeneity investigation (serum samples only)

Before pooling all data sets together, we investigated whether selected variables had any impact on overall test accuracy by adding covariates for each variable of interest to a regression model (*Table 14*). Bearing in mind that only 17 studies were included in these analyses, thereby reducing the overall power to detect any effect, the only variables to reach statistical significance were study design clearly reported as prospective (RDOR 0.14; 95% CI 0.03 to 0.66) and reference test clearly reported as interpreted blinded (RDOR 0.10; 95% CI 0.02 to 0.41), that is, both groups were found to have significantly lower accuracy than when these factors were not clearly reported.

The commercial tests as a group were not significantly more or less accurate than the in-house test group.

SROC analysis

Serum samples

The overall pooled analysis indicates that the anti-TB antibody tests are not very accurate at detecting the presence of TB infection (*Table 15*), DOR 6.93 (95% CI: 2.24 to 21.49); associated

TABLE 14 *Anti-TB antibody test evaluations in pulmonary TB – regression analyses to identify source(s) of heterogeneity (serum samples only)*

Comparison	*n* (17)	Model parameters (95% CI) and *p*-value		
		Coefficient	*p*-Value	RDOR[a]
Test type				
Commercial vs in-house	8/9	0.95	0.29	2.57 (0.41 to 15.99)
Institute				
Primary hospital vs referral hospital	7/9	1.14	0.17	3.13 (0.57 to 17.28)
Reference standard				
Culture plus clinical vs culture alone	8/3	–1.12	0.37	0.32 (0.02 to 4.44)
Culture plus other vs culture alone	4/3	–2.48	0.06	0.08 (0.01 to 1.18)
No culture vs culture alone	2/3	–0.26	0.86	0.77 (0.03 to 17.59)
Design				
Prospective vs unknown	10/7	–1.95	0.02	0.14 (0.03 to 0.66)
Quality factors				
Patients *not* representative vs unknown	11/6	–0.90	0.32	0.41 (0.06 to 2.62)
Index test blinded vs unclear	8/9	–1.35	0.12	0.26 (0.05 to 1.48)
Reference test blinded vs unclear	6/11	–2.32	0.00	0.10 (0.02 to 0.41)
TB prevalence				
Prevalence <10% vs prevalence 20+%	1/9	–1.48	0.41	0.23 (0.01 to 9.57)
Prevalence 10–20% vs prevalence 20+%	7/9	–1.28	0.16	0.28 (0.04 to 1.76)

[a] Relative diagnostic odds ratio, i.e. in one group compared with the other.

sensitivity in particular being very low at 48.3%. The commercial tests did perform slightly better than the in-house tests but the difference was very slight (DOR 7.30 compared with 3.88).

Results when each of the five design-related criteria were individually applied is reported in *Table 16*. Accuracy was generally reduced in each higher quality subgroup, except in the group of studies that were either prospective in design and/or used blinded index test interpretation (DOR 11.96; 95% CI: 11.96 to 59.58, sensitivity still low at 50.4%).

When more than one of the design-related criteria was applied together, the overall accuracy progressively dropped. The DOR in the eight studies meeting two criteria (hospital based and combined reference standard used) was 5.42, and in the single study of Detect TB that met all five criteria it was 2.60.

Non-serum samples

The anti-TB antibody tests overall performed better when applied to non-serum samples (*Table 17*). The DOR in the three data sets in respiratory specimens was 17.05 (95% CI: 0.07 to 4187.22) and in the single data set using urine specimens (DOR 18.83; 95% CI: 12.92 to 27.46). None of the studies met the first two design-related criteria.

Summary

The anti-TB antibody tests were found not to perform well with either serum samples or respiratory specimens. When all studies were included there was some indication that some of them were performing reasonably well (sensitivity and specificity pairs in the upper left-hand quadrant of the ROC plot); however, all of these studies were excluded when the quality criteria were applied.

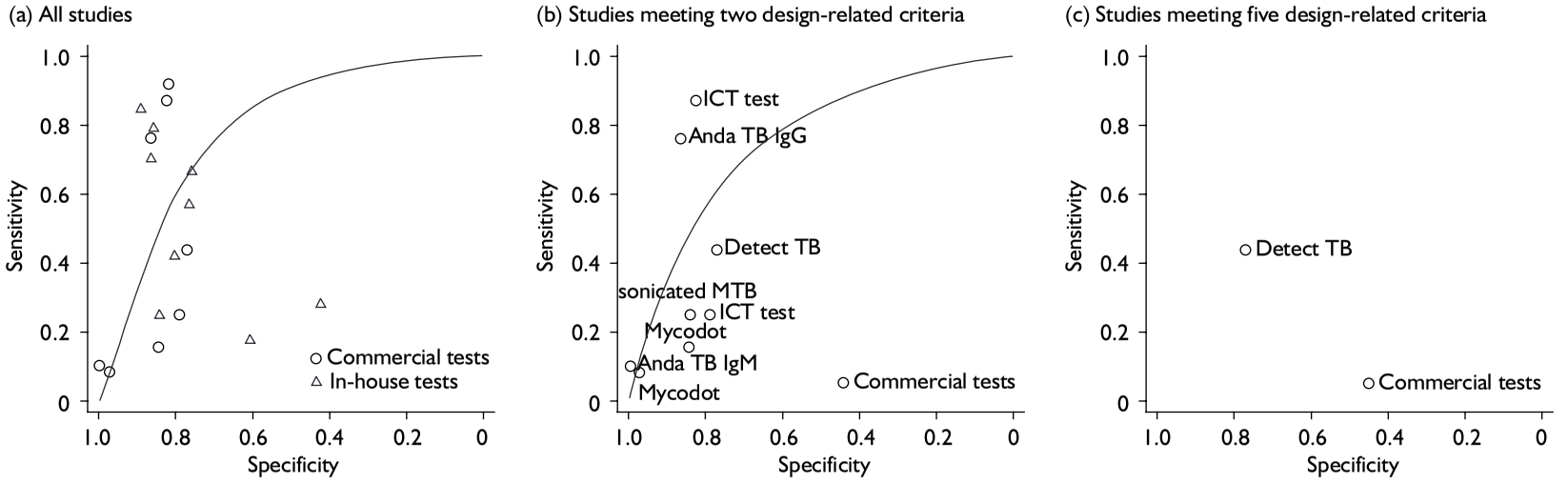
The only factors identified that explained the heterogeneity of the studies were prospective design and reference test blinding, both of which led to lower accuracy. The number of studies in this analysis was small, however, and it is not surprising that other factors were not found to have a significant effect.

Other biochemical tests

Description and quality of included studies

We included six data sets relating to other tests that have been used to identify pulmonary tuberculosis. Summary details per study are provided in Appendix 10.

TABLE 15 *Anti-TB antibody test evaluations in pulmonary TB (serum samples only) – SROC models and associated SROC curves*



	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
All data sets	**17**	**6.93 (2.24 to 21.49)**	**0.42**	**48.3**	**84.8**	**8**	**5.42 (0.68 to 43.28)**	**0.86**	**33.1**	**90.6**	**1**	**2.60 (1.40 to 4.82)**	**NA**	**43.8**	**76.9**
By test															
Commercial	8	7.30 (1.95 to 27.24)	NA	87.9	50.0	7	6.35 (0.59 to 67.98)	0.85	34.4	91.3	1	2.60 (1.40 to 4.82)	NA	43.8	76.9
In-house	9	3.88 (1.01 to 15.00)	NA	52.7	77.7	1	1.77 (0.44 to 7.10)	NA	25.0	84.1					

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

TABLE 16 *Anti-TB antibody test evaluations in pulmonary TB (serum samples only) – effect of individual quality criteria on test accuracy*

	n[a]	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
All tests	**17**	**6.93 (2.24 to 21.49)**	**0.42**	**48.3**	**84.8**
Hospital-based	16	6.11 (1.84 to 20.36)	0.51	46.1	84.8
Culture + clinical reference test	8	5.42 (0.68 to 43.28)	0.86	33.1	90.6
Index test blinded/study prospective	11	11.96 (2.40 to 59.58)	0.06	50.4	77.4
Reference test blinded	6	0.57 (0.06 to 5.08)	0.35	23.4	78.9
Patients representative	11	3.52 (0.70 to 17.65)	0.94	39.4	84.9

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

Adenosine deaminase tests

The two ADA tests were evaluated in the same study[200] in serum samples from 110 patients. The prevalence of TB in the sample was 59%. The study was prospective and hospital based with a representative sample of patients. Index and reference test blinding was not reported and a combined reference strategy of culture plus clinical diagnosis was used.

Antigen test

The antigen test study[201] evaluated the use of antibodies targeted against LAM and H37Rv antibodies in respiratory samples from 62 patients. TB prevalence was 39%. The patient sample was judged to be representative but no information on blinding or study design could be derived. The reference standard used was culture alone.

Cytokine tests

The two cytokine tests were evaluated in the same study[190] in pleural fluid samples from 14 patients. The prevalence of TB in the sample was 36%. The study was hospital based with a representative sample of patients. Index and reference test blinding and study design were reported. The reference standard used was culture alone.

Other tests

One study[202] evaluated the use of tuberculostearic acid (TBSA) to detect TB in respiratory samples from 39 patients. The prevalence of TB in the sample was 54%. The study was hospital based and prospective with a representative sample of patients. The index test was reported to be interpreted blinded, but could not be judged for the reference test. The reference standard used was culture plus histology and X-ray.

Results

The plot of sensitivity against specificity for each group of tests is presented in *Figure 2(f–i)*.

Adenosine deaminase tests

Both of the ADA tests perform very poorly for detection of TB (*Table 18*), with a pooled DOR of only 1.25.

Antigen test

The antigen test evaluated had relatively high specificity (92.1%) but sensitivity was poor at only 62.5%) (*Table 19*).

Cytokine tests

One of the cytokine tests (assessing interferon-γ levels in pleural fluid) performed relatively well, with a sensitivity of 91.7% and a specificity of 85.0%, but the other (assessing interleukin levels) performed poorly (*Table 20*).

Other tests

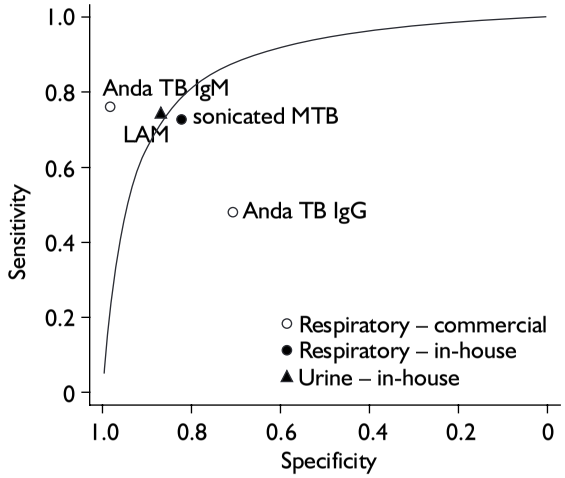
Based on one study, TBSA does not appear to be a sensitive indicator of the presence of TB infection (*Table 21*).

Summary

There is no indication that any of these other tests have a role in the diagnosis of pulmonary TB.

TABLE 17 *Anti-TB antibody test evaluations in pulmonary TB (serum samples only) – SROC models and associated SROC curves*

(a) All studies[a]



(b) Studies meeting two design-related criteria

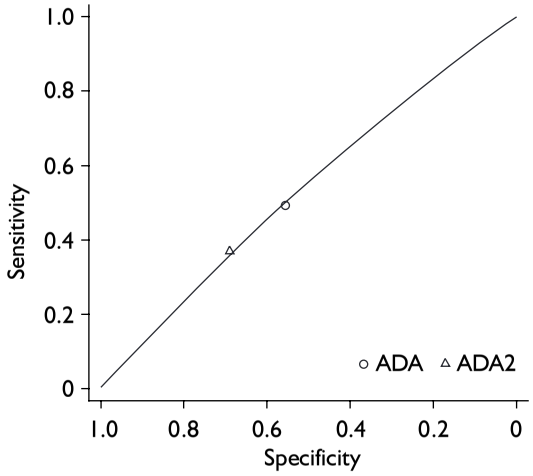
None

	n[b]	DOR (95% CI)[c]	Test of symmetry[d]	Se[e] (%)	Sp[e] (%)
Respiratory specimens	3	17.05 (0.07 to 4187.22)	NA	66.5	89.6
Urine	1	18.83 (12.92 to 27.46)		74.0	86.9

[a] SROC curve relates only to respiratory samples.
[b] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[c] DOR where sensitivity = specificity.
[d] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[e] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

TABLE 18 *Adenosine deaminase tests evaluations in pulmonary TB – SROC model and SROC curve*

(a) All studies



(b) Studies meeting two design-related criteria

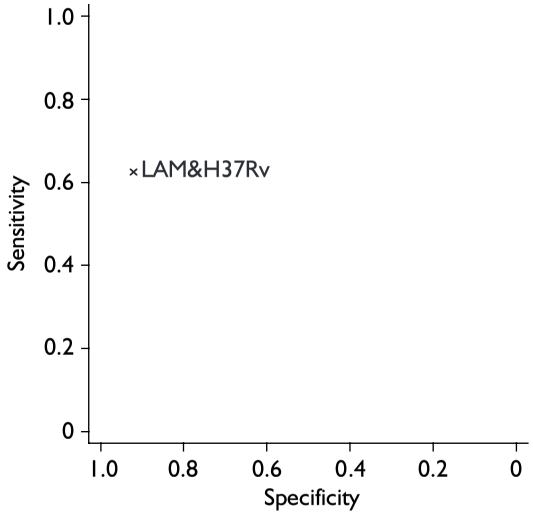
None

Comparison	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
ADA (both serum samples)	2	1.25 (0.82 to 1.92)	NA	43.0	62.5

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

TABLE 19 *Antigen test evaluations in pulmonary TB – study results*

(a) All studies



(b) Studies meeting two design-related criteria

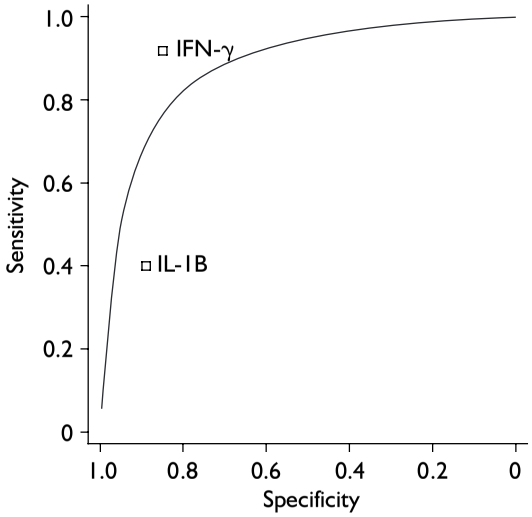
None

Comparison	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
LAM and H37Rv (respiratory sample)	1	19.44 (4.61 to 82.06)	NA	62.5	92.1

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not, if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

TABLE 20 *Cytokine test evaluations in pulmonary TB – SROC models*

(a) All studies



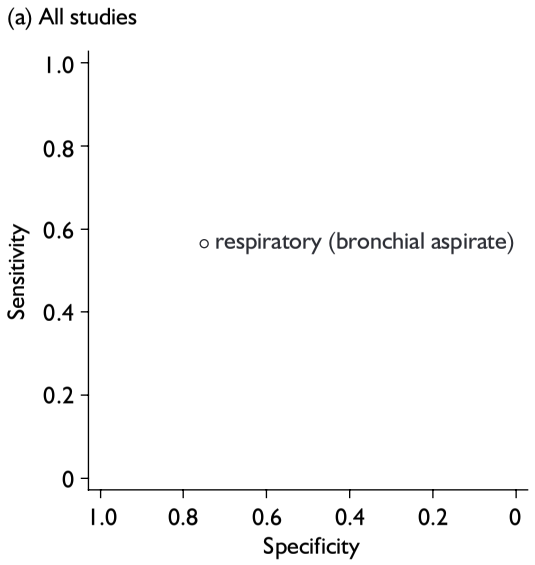
(b) Studies meeting two design-related criteria

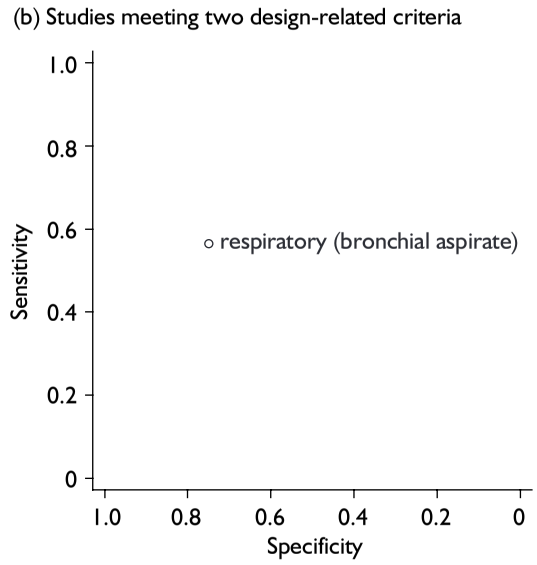
None

Comparison	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
All (both PF samples)	**2**	**18.23 (0.00 to >1,000,000)**	**NA**	**73.0**	**87.1**
Interferon-γ	1	62.33 (2.13 to 1822.74)	NA	91.7	85.0
Interleukin	1	5.33 (0.34 to 82.83)	NA	40.0	88.9

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

TABLE 21 *Other miscellaneous test evaluations in pulmonary TB – SROC models*





Comparison	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
TBSA (respiratory sample)	1	3.9 (0.96 to 15.82)		56.5	75.0	1	3.9 (0.96 to 15.82)		56.5	75.0

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

TABLE 22 *Pulmonary TB – comparisons between test types*[a]

		NAAT (*n* = 106)			**NAAT plus molecular probe (*n* = 4)**			**Phage (*n* = 5)**	
NAAT + molecular probe	106/4	ROR = 2.74, $p = 0.29$	←						
	1/1	–	–						
Phage	106/5	ROR = 0.59, $p = 0.54$	↑	4/5	ROR = 0.27, $p = 0.31$	↑			
	2/2	*ROR = 0.22*, p = *0.44*	↑	0/0	–	–			
Adenosine deaminase	106/2	ROR = 0.01, $p < 0.01$	↑↑	4/2	ROR = 0.003, $p = 0.01$	↑↑	5/2	ROR = 0.01, $p = 0.01$	↑↑
	0/0	–	–	0/0	–	–	0/0	–	–
Anti-TB antibody	106/21	ROR = 0.04, $p < 0.01$	↑↑	4/21	ROR = 0.01, $p < 0.01$	↑↑	5/21	ROR = 0.05, $p < 0.01$	↑↑
	2/1	*ROR = 0.005*, p = *0.02*	↑↑	0/0	–	–	0/0	–	–
Antigen	106/1	ROR = 0.12, $p = 0.26$	↑	4/1	ROR = 0.04, $p = 0.19$	↑	5/1	ROR = 0.18, $p = 0.29$	↑
	0/0	–	–	0/0	–	–	0/0	–	–
Cytokine	106/2	ROR = 0.14, $p = 0.13$	↑	4/2	ROR = 0.04, $p = 0.09$	↑	5/2	ROR = 0.17, $p = 0.19$	↑
	1/2	*ROR = 0.06*, p = *0.41*	↑	0/0	–	–	0/0	–	–
		Adenosine deaminase (*n* = 2)			**Anti-TB antibody (*n* = 21)**			**Antigen (*n* = 1)**	
Anti-TB antibody	2/21	ROR = 5.82, $p = 0.18$	←						
	0/0	–	–						
Antigen	2/1	ROR = 15.51, $p = 0.01$	←←	21/1	ROR = 3.04, $p = 0.54$	←			
	0/0	–	–	0/0	–	–			
Cytokine	2/2	ROR = 14.55, $p = 0.16$	←	21/2	ROR = 2.46, $p = 0.49$	←	1/2	ROR = 0.94, $p = 0.98$	↑
	0/0	–	–	0/0	–	–	0/0	–	–

ROR, relative odds ratio – ratio of DORs.

[a] ↑, ←, arrows point towards test with higher DOR. Upper cells indicate unpaired comparisons; lower cells denoted in italics indicate results from paired comparisons.

Empirical comparisons between test types

NAAT test comparisons

The result of the comparisons between different types of test (*Table 22*) show that the NAAT tests are superior to all of the other tests analysed, except in comparison with the simultaneous combination of molecular amplification with molecular probe, where the DOR was almost three times that of the NAAT group, although the results was not significant [ratio of odds ratios (ROR) 2.74, $p = 0.29$]. In the one study with paired data (i.e. both tests evaluated in the same study),[203] the DOR for the AMTD test was 872 compared with 5056 for the BD ProbeTec ET test, the former being due to a lower sensitivity estimate (85% compared with 94% for the combined amplification with probe).

For most of the other NAAT test comparisons in the unpaired comparisons the NAATs were at least eight times more accurate than the other tests and this was statistically significant for the comparisons with ADA tests and with anti-TB antibody tests. The phage-based tests, although highly specific, were variable in terms of sensitivity, and the overall DOR was 0.6 times that of the molecular amplification, $p = 0.54$. The two studies with paired data[196,198] suggest in fact that the DOR of the phage tests is only 0.22 times that of the NAAT tests, although again the result is not significant. The much lower DOR for the phage tests is due to an extremely low sensitivity estimate (27%) in one study.[198]

Other test comparisons

A similar pattern of results can be seen for the other comparisons (*Table 22*). The amplification plus molecular probe tests are better than all of the other tests evaluated, but significantly so only for the comparison with ADA tests and anti-TB antibody tests.

The phage tests are better than all other tests except for NAAT tests and amplification plus molecular probe tests.

Summary

Overall ranking of tests appears to be amplification plus molecular probe tests, NAAT tests, phage tests. None of the serodiagnostic or biochemical tests show any promise for diagnosis of pulmonary TB.

Discussion: test accuracy in pulmonary TB infection

The tests showing the most promise for diagnosis of pulmonary TB are NAAT tests, amplification plus molecular probe tests and possibly phage tests. However, the overall quality and reporting of studies were poor, and were similar across all test types, making the true accuracy of the tests difficult to estimate with certainty, especially when small numbers of studies are available. For example, in over 60% of cases it was impossible to determine even whether the study design was prospective or retrospective and whether blinded test interpretation had been used. Less than half of studies used our designated 'ideal' reference standard of culture plus high clinical suspicion with or without other diagnostic interventions.

Patients were judged to be representative in more than 50% of studies but we only required the study to state that patients had been 'suspected' of pulmonary TB; no further information on the demographics, symptoms or clinical features of disease was required and, in many cases, such information was in fact not provided. Sample sizes for the main tests evaluated were at least generally fairly large with means of 213 and 606, although in each case the SDs were also large. The mean prevalence of TB ranged from 29 to 32%.

There is considerable evidence to support the use of NAAT tests for the diagnosis of pulmonary TB; in particular, the AMTD test appears to perform better than other currently available commercial tests. However, accuracy was lower in the better designed and reported studies, reducing the sensitivity estimate to only 79%. Sensitivity of the higher quality in-house tests was better (90%), but specificity was sacrificed considerably (78%). Although the percentage of smear-positive patients included in the studies did not significantly affect accuracy in the heterogeneity investigations, there was a trend towards higher accuracy in studies with higher prevalences of smear-positive patients and indeed the subgroup analyses according to smear status showed considerably lower accuracy in smear-negative patients, especially in studies using a combined reference standard (sensitivity 46%, specificity 98% based on three studies of the Amplicor test).

The main result for smear-negative patients from our SROC model (DOR 35.80, 95% CI: 17.68 to 72.51) was much lower than that obtained in another recently published meta-analysis,[188] which produced a pooled DOR of 51.11 (95% CI: 27.56 to

94.78) from 16 studies giving per patient data. However, the estimates of sensitivity and specificity resulting from our model (sensitivity 73.4%, specificity 93.7%) are very similar to those of Sarmiento and colleagues,[188] who directly pooled sensitivity and specificity estimates using random effects methods to give a pooled sensitivity of 72% and a specificity of 96%. They concluded that PCR should not be routinely used for diagnosis of smear-negative pulmonary TB infection. Our results suggest that NAAT tests can be used to rule in disease (i.e. a positive NAAT test result is unlikely to be a false negative); however, due to low sensitivity cannot be used to rule out disease. Currently, both Amplicor and the AMTD tests are FDA approved for use in smear-positive and -negative patients.

In-house NAAT tests performed fairly well, but are only really relevant to the NHS if they can be shown as reliable as, but less expensive than, the commercial tests. A laboratory not currently providing a service such as NAAT testing will be primarily interested in cost, accuracy and quality assurance.

The four studies that combined amplification with a molecular probe test suggest that this approach may in fact be superior to using molecular amplification alone, but none of these studies were hospital based or used a combined reference standard, so the true accuracy cannot yet be determined. The phage tests have been shown to have high specificity in the five available studies, but again it is not yet clear whether these tests have sufficiently high sensitivity in smear-negative samples to recommend their routine use in practice. The main reasons for the observed variation in sensitivity are likely to be biological, relating to the stability of the phage and the interaction of two living biological systems (the phage and the mycobacteria). Sensitivity in smear-positive patients has so far been shown to be between 80 and 90%. Phage tests have some potential advantages over NAAT tests because they detect only viable mycobacteria and can yield antibiotic sensitivities within 2–3 days.

None of the other tests evaluated in pulmonary TB show any real promise for diagnosis.

Chapter 6

Results: detection of miscellaneous extra-pulmonary tuberculosis infection

We identified 30 test comparisons, 22 evaluating NAAT tests, six anti-TB antibody tests and one each molecular probe tests and ADA tests (*Table 23*). ROC plots of each sensitivity and specificity pair for each group of tests are provided in *Figure 7(a–d)*.

Nucleic acid amplification tests

Description and quality of included studies

A summary of key characteristics across all NAAT data sets is given in *Table 24*, with details per study provided in Appendix 11 grouped according to evaluation of commercial or in-house tests.

Tests

More than half (13/22) of the evaluations related to commercial NAAT tests, including seven using the the Gen-Probe Amplified Mycobacterium Direct Test (AMTD), four of the Roche Amplicor MTB test, and one each of the Abbott Laboratories LCx test and the Amplicis Myco B test.

Of the nine in-house test evaluations, six used the IS6110 target sequence. The other studies employed other gene targets: the 65-kDa (two studies) and 16SrRNA and IS986 (one data set each).

TABLE 23 *Miscellaneous extrapulmonary TB samples – summary of data sets identified*

	All
Total	30
NAAT tests	**22**
Commercial tests	13
AMTD	7
Amplicor	4
LCx	1
Amplicis (Myco B)	1
In-house tests	9
IS6110	6
Other targets	3
Molecular probe tests	**1**
LiPA	1
Anti-TB antibody tests	**6**
Commercial	4
Anda TB IgG	1
Anda TB IgM	1
ICT	2
In-house	2
H37Rv	1
Sonicated MTB (unspecified)	1
Adenosine deaminase tests	**1**
ADA	1

TABLE 24 *NAAT test evaluations in miscellaneous extrapulmonary TB – summary of key study characteristics*

	All samples
Total no. of studies	22
Mean sample size	205 (SD 231; range 23–1090)
Commercial	13
Mean sample size	269 (SD 279; range 23–1090)
In-house	9
Mean sample size	113 (SD 88; range 34–308)
Reference standard	
Culture + clinical diagnosis ± other	10 (45%)
Culture + other	3 (14%)
Culture alone	9 (41%)
Clinical diagnosis alone	0
Disease prevalence (mean, SD, range)	22% (SD 25; range 3–59%)
Setting	
Hospital-based	12 (55%)
Laboratory-based	9 (41%)
Unknown	1 (5%)
Sample type	
>20% sputum samples	2 (9%)
No sputum samples included	18 (91%)
Patients representative?	
Yes	14 (64%)
No	3 (14%)
Unclear	5 (23%)
Study design prospective?	
Yes	2 (9%)
No	2 (9%)
Unclear	18 (91%)
Index test interpreted blinded?	
Yes	5 (23%)
No	0
Unclear	17 (77%)
Reference test interpreted blinded?	
Yes	5 (23%)
No	1 (5%)
Unclear	16 (73%)

Reference tests
Nine of the 22 data sets (41%) used culture alone as the reference standard, 10 (45%) combined culture with clinical symptoms (with or without an assessment of response to anti-TB therapy or chest X-ray) and three used culture plus histology or X-ray.

Sample details
Twelve data sets were from hospital-based studies, that is, recruited referred patients, and nine from laboratory-based studies (recruiting samples rather than patients). In the remaining study, the source of the population was not clear. Twenty studies did not include any sputum samples; in two less than 20% of samples were sputum (these datsets were included in this section as the samples studies were predominantly extrapulmonary).

Across all 22 data sets, the mean number of patients recruited was 205 (SD 231; range 23–1090). The mean was much lower for the studies of in-house tests (113, SD 88) compared with commercial tests (n = 269, SD 279). Mean prevalence of TB was 22% (SD 25%).

In 64% (14/22) of studies we judged the patient sample to have been representative (i.e. the study at least stated that that patients 'suspected' of having TB were recruited). The sample was judged to be unrepresentative in three studies and was unclear in five.

Test interpretation
Index test interpretation was reported as blinded in 5/22 studies and could not be determined in the rest. Two studies were clearly prospective in design and two were retrospective. The reference standard was clearly reported as interpreted blinded in five studies, three of which also reported blinded index test interpretation.

Results
Sensitivity and specificity
The plot of sensitivity against specificity for each study is presented in *Figure 7(a)*, and according to test and/or target sequence in *Figure 8*. There is a considerable range in both sensitivity and specificity estimates, for both commercial and in-house tests, although visually specificity varies less for the commercial tests. The heterogeneity tests reported in *Table 25* are all statistically significant ($p < 0.01$) except for sensitivity of the AMTD test. The ROC plot of the commercial group by test (*Figure 8a*) confirms that the AMTD data sets are predominantly clustered together in the top left-hand corner of the plot.

Heterogeneity investigation
The results of the heterogeneity investigations (*Table 26*) demonstrate that none of the commercial tests are significantly more or less accurate than the in-house tests and, for the in-house tests, the use of the target sequence IS6110 has no significant impact compared with the other target sequences used.

The only two variables to reach statistical significance were prospective design reported (accuracy 17 times higher than in retrospective/unknown design, $p = 0.02$) and blinded reference test interpretation (accuracy significantly lower than not blinded or blinding not reported, $p = 0.01$). Accuracy was also higher in laboratory-based studies and in studies with higher prevalence, but this did not reach statistical significance.

SROC analysis
The overall pooled analysis produces a DOR of 54.22 (95% CI: 16.06 to 183.03) with an associated sensitivity of 78.1% and a specificity of 95.8% (*Table 27*). Accuracy appears slightly better for the commercial tests than the in-house tests in terms of DORs; although they have lower sensitivity, the specificity is much higher. Individually, the AMTD test has the highest summary sensitivity and specificity (85.6 and 97.7%, respectively) (*Figure 9*).

When studies were limited to hospital based and those using a combined reference standard, sensitivity dropped considerably (to 63%) and specificity also fell (to 89%). Individually, the AMTD test still had the best performance with sensitivity and specificity of 88 and 97%, respectively.

Summary
NAAT test performance is clearly highly variable in extrapulmonary samples, but our results indicate that it potentially has a role to play in the diagnosis of extrapulmonary TB.

Other tests
Amplification with simultaneous molecular probe tests
One study of the Inno-LiPA test in mixed extrapulmonary specimens from 75 patients was included[204] (Appendix 12). It was a hospital-based study using a combined reference standard of culture plus clinical TB diagnosis. Index test interpretation was apparently blinded (amplification and probe tests were carried

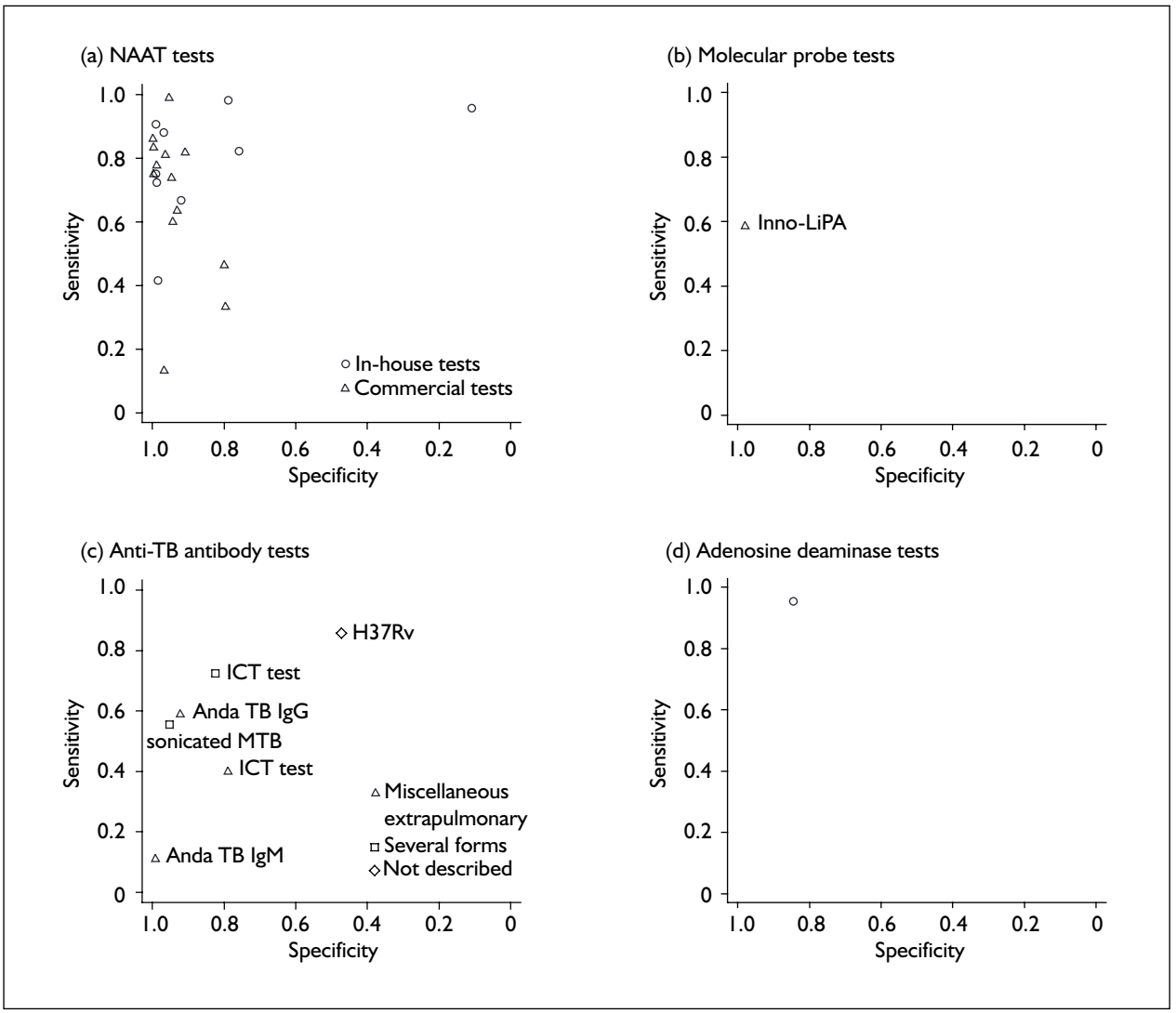


FIGURE 7 *Miscellaneous extra-pulmonary TB studies – ROC plots*

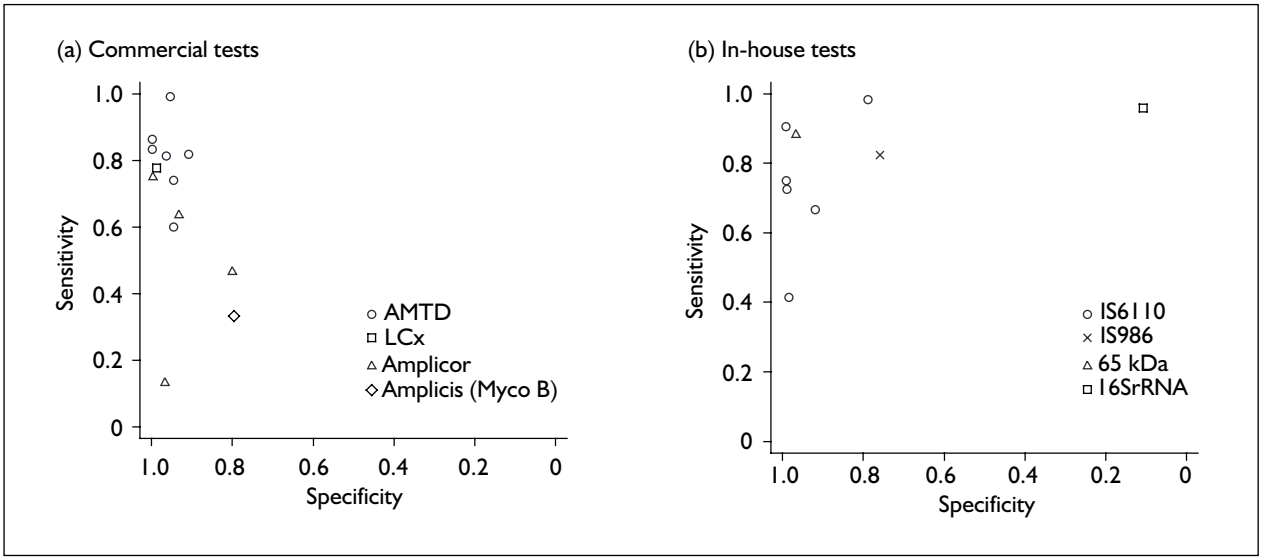


FIGURE 8 *NAAT tests in miscellaneous extrapulmonary TB samples – ROC plots by type of test*

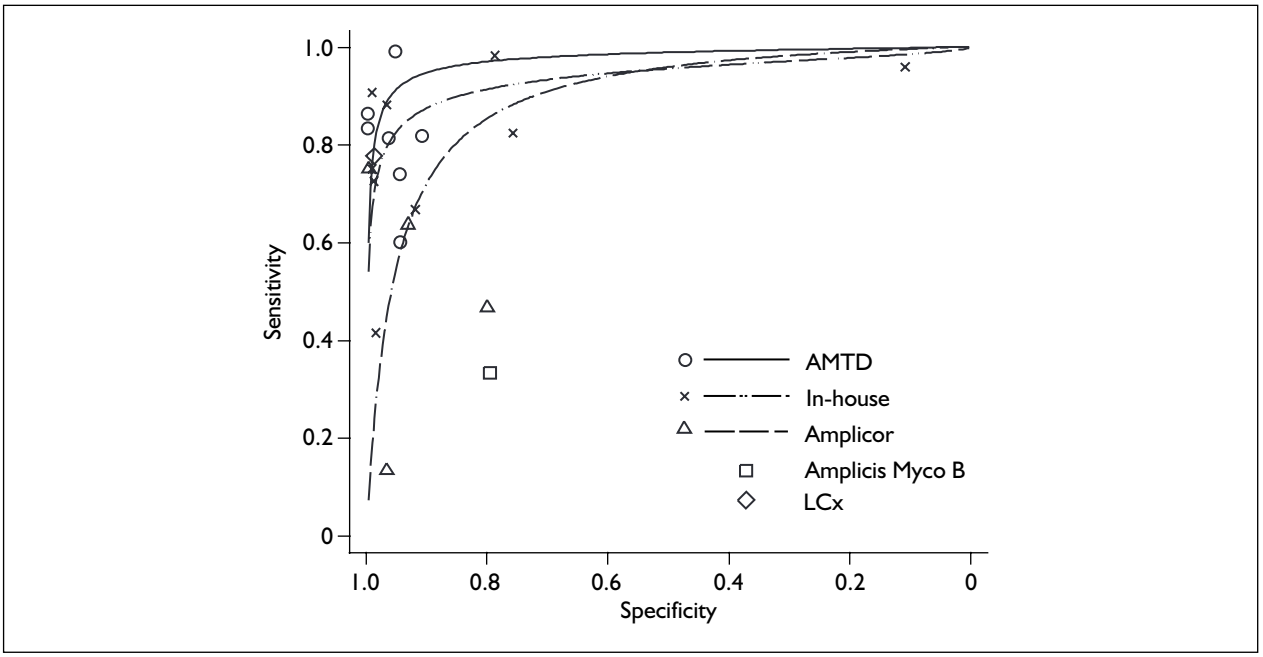


FIGURE 9 *NAAT tests in extrapulmonary TB – separate SROC regression curves per subgroup of studies*

TABLE 25 *Miscellaneous extrapulmonary TB samples – tests for heterogeneity of sensitivities and specificities*

		Sensitivity		Specificity	
Comparison	***n***	***Q statistic***	***p*-Value**	***Q statistic***	***p*-Value**
All NAAT comparisons	**22**	**104**	**<0.01**	**247**	**<0.01**
Commercial tests	13	61	<0.01	104	<0.01
AMTD	7	11	0.10	16	0.01
Amplicor	4	21	<0.01	58	<0.01
LCx	1				
Amplicis (Myco B)	1				
In-house tests	9	43	<0.01	74	<0.01
IS6110	6	37	<0.01	30	<0.01
Other targets	3	1	0.59	33	<0.01
Anti-TB antibody tests					
All data sets	6	75	<0.01	78	<0.01

out at another site with final diagnoses established at a later date). TB prevalence was 68%. The sensitivity of the test was found to be 59% and the specificity 100%.

Anti-TB antibody tests

Four studies provided six data sets evaluating anti-TB antibody tests, four of commercial tests and two of the in-house tests (Appendix 12). All evaluations were carried out with serum samples. A combined reference standard was used in three of the four commercial test evaluations. Two studies reported using blinded index test interpretation, one of which also reported blinded reference test interpretation. One study was prospective in design and in two we judged the patient sample to have been representative.

The SROC results presented in *Table 28* show that none of the tests performed well. Although specificity was reasonably high for the commercial test evaluations (92.5%), sensitivity was less than 50%.

Adenosine deaminase tests

Makhlouf and colleagues[205] evaluated ADA in 90 patients. The reference standard used was culture plus histology, and none of the study design-related criteria could be determined from the study report. Sensitivity was 100% and specificity 85%.[205]

Summary

Insufficient data are available on which to base any recommendations regarding these studies in the evaluation of extrapulmonary TB.

TABLE 26 *Miscellaneous extrapulmonary TB samples – regression analyses to identify source(s) of heterogeneity*

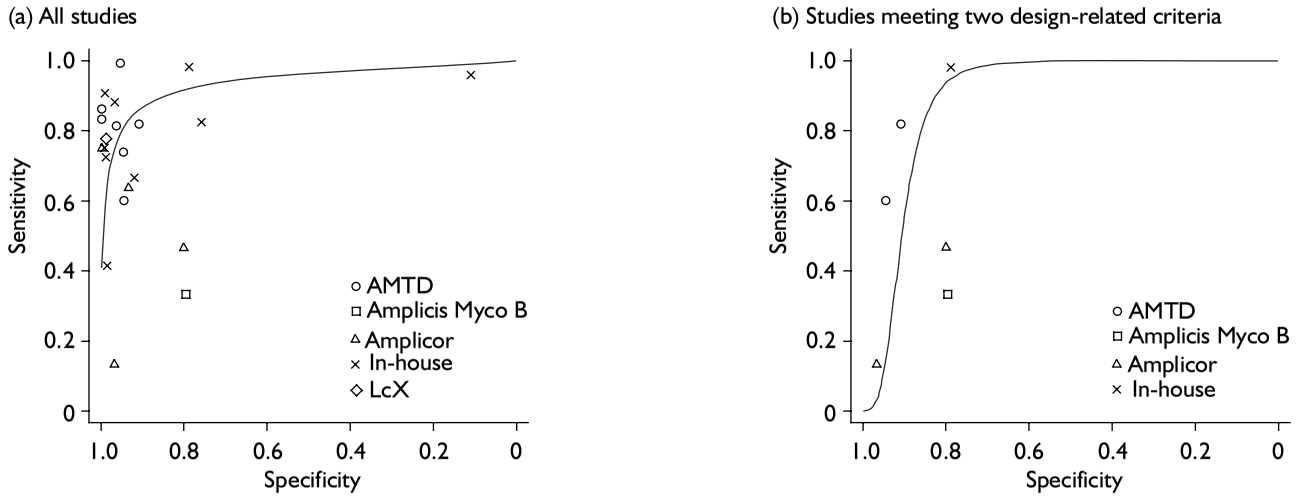
Comparison	*n* (22)	Model parameters (95% CI) and *p*-value		
		Coefficient	*p*-Value	RDOR[a]
Test type				
AMTD vs in-house	7/9	0.86	0.40	2.36 (0.28 to 19.70)
Amplicor vs in-house	4/9	–1.93	0.14	0.14 (0.01 to 2.04)
LcX vs in-house	1/9	0.59	0.78	1.81 (0.02 to 152.59)
Amplicis Myco B vs in-house	1/9	–4.03	0.07	0.02 (0.00 to 1.42)
For in-house only: IS6110 vs other targets	6/3	1.23	0.42	3.42 (0.11 to 107.11)
Institute				
Laboratory-based vs hospital-based	9/13	1.59	0.10	4.88 (0.74 to 32.37)
Reference standard				
Culture plus clinical vs culture alone	9/12	0.26	0.79	1.29 (0.18 to 9.53)
Culture plus other vs culture alone	3/12	2.86	0.06	17.42 (0.89 to 342.64)
Design				
Prospective vs retrospective/unknown	2/20	3.63	0.02	37.54 (1.72 to 821.22)
Quality factors				
Patients representative vs unrepresentative/unknown	14/9	0.56	0.63	1.76 (0.16 to 18.96)
Index test blinded vs not blinded/unknown	5/17	–1.24	0.28	0.29 (0.03 to 2.96)
Reference test blinded vs not blinded/unknown	5/17	–2.92	0.01	0.05 (0.01 to 0.38)
TB prevalence				
Prevalence <10% vs prevalence 30+%	2/6	–1.52	0.43	0.22 (0.00 to 11.48)
Prevalence 10–20% vs prevalence 30+%	5/6	–1.23	0.39	0.29 (0.02 to 5.48)
Prevalence 20–30% vs prevalence 30+%	9/6	–0.13	0.91	0.88 (0.07 to 11.03)
% patients smear positive				
<5% smear positive vs >30%	3/2	4.33	0.02	75.95 (2.07 to 2791.17)
5–10% smear positive vs >30%	3/2	0.99	0.45	2.68 (0.16 to 45.28)
10–20% smear positive vs >30%	5/2	2.46	0.09	11.71 (0.65 to 211.23)
20–30% smear positive vs >30%	2/2	3.97	0.02	52.78 (2.60 to 1070.79)

[a] Relative diagnostic odds ratio, i.e. in one group compared with the other.

Discussion: test accuracy in miscellaneous extrapulmonary samples

The NAAT tests are the only group of tests for which there is a reasonable amount of evidence in mixed extrapulmonary samples. Accuracy varies considerably, but a number of studies demonstrating high specificity and relatively high sensitivity are available. However, there are a number of quality concerns regarding these data. A high proportion of the studies (41%) used culture alone, known to perform poorly with fluid samples, as the reference standard. Blinded test interpretation was not reported in over 70% of studies and the study design was unknown in 91%.

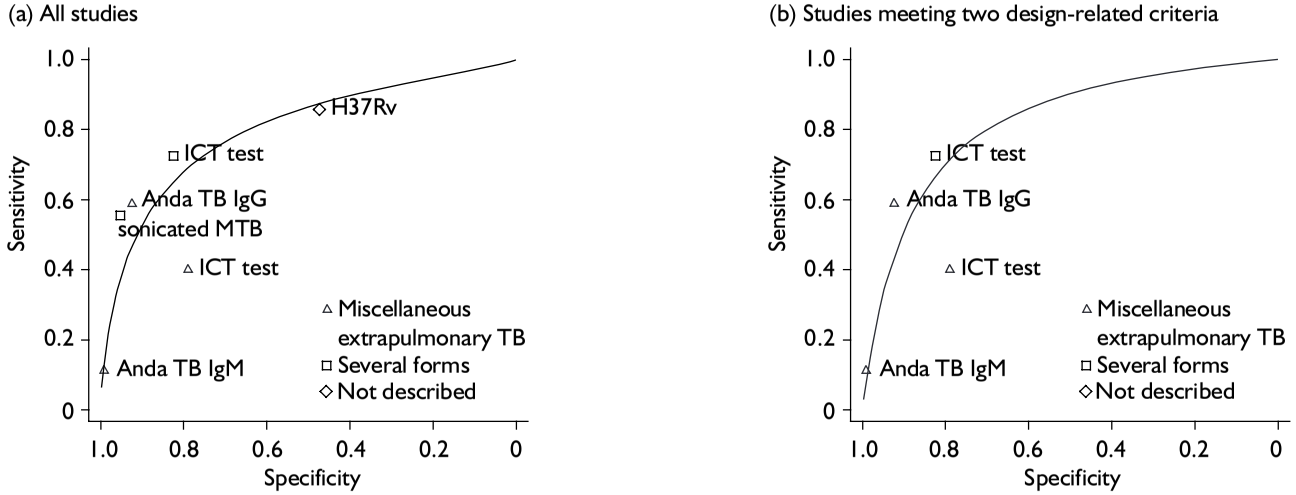
TABLE 27 *Miscellaneous extrapulmonary TB samples – SROC regression models and associated SROC curve*



	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
All tests	22	54.22 (16.06 to 183.03)	0.26	78.1	95.8	6	30.42 (4.79 to 193.11)	0.10	63.0	89.0
Commercial	**13**	**71.25 (4.54 to 1118.76)**	**0.88**	**72.6**	**96.9**	**5**	**15.65 (0.32 to 771.00)**	**0.57**	**46.1**	**90.4**
AMTD	7	181.20 (7.92 to 4146.43)	0.72	85.6	97.7	2	33.65 (0.99 to 1140.41)	NA	88.2	97.4
Amplicor	4	23.27 (0.44 to 1242.39	NA	47.8	96.2	2	3.95 (0.85 to 18.47)	NA	26.8	91.5
LcX	1	272.35 (102.83 to 721.32)	NA	77.6	98.7	0				
Amplicis Myco B	1	1.95 (0.16 to 23.73)	NA	33.3	79.6	1	1.95 (0.16 to 23.73)	NA	33.3	79.6
In-house	**9**	**57.70 (15.09 to 220.66)**	**0.11**	**84.6**	**93.5**	**1**	**200.57 (47.54 to 846.22)**		**98.0**	**79.0**
Subgroups meeting each design-related criterion										
Hospital-based	13	28.30 (5.28 to 151.68)	0.35	73.2	93.9					
Combined reference test	10	122.44 (13.81 to 1085.23)	0.27	77.1	94.9					
Index test blinded	7	126.98 (3.42 to 4708.88)	0.81	81.1	95.9					
Reference test blinded	5	8.43 (0.06 to 1107.08)	0.91	49.1	91.1					
Patients representative	14	34.54 (3.03 to 394.44)	0.23	72.9	91.1					

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

TABLE 28 *Anti-TB antibody test evaluations in miscellaneous extrapulmonary TB – SROC model and SROC curve*



Comparison	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
All data sets	**6**	**7.88 (2.26 to 27.43)**	**NA**	**53.6**	**89.7**	**4**	**9.30 (2.27 to 38.18)**	**NA**	**42.8**	**92.5**
Commercial	4	9.30 (2.27 to 38.18)	NA	42.8	92.5	4	9.30 (2.27 to 38.18)	NA	42.8	92.5
In-house	2	11.60 (0.00 to 201,005.99)	NA	73.3	80.9	0				

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

Chapter 7

Results: detection of pleural tuberculosis infection

We identified 97 test comparisons in pleural TB: 76 in pleural fluid specimens, 11 in serum samples and 10 using pleural/serum ratio (see *Table 29*). ROC plots of each sensitivity and specificity pair for each group of tests are provided in *Figure 10(a–g)*.

Nucleic acid amplification tests

Description and quality of included studies

We included 20 data sets relating to NAATs, all of which used pleural fluid specimens. These

TABLE 29 *Pleural TB – summary of data sets identified*

	All	PF	Serum	Other
TOTAL	100	79	11	10
NAAT tests	**20**	**20**	**0**	**0**
Commercial	6	6	0	0
AMTD	3	3	0	0
Amplicor	2	2	0	0
Amplicis (Myco B)	1	1	0	0
In-house	14	14	0	0
IS6110	9	9	0	0
Other targets	5	5	0	0
Adenosine deaminase tests	**42**	**36**	**2**	**4 p/s**
ADA	31	27	2	2
ADA2	1	1	0	0
ADA/ADA1	1	1	0	0
ADA/lysozyme	1	0	0	1
ADA/total protein	1	1	0	0
2′-Deoxyadenosine deaminase	1	1	0	0
Anti-TB antibody tests	**7**	**4**	**3**	**0**
Commercial	6	3	3	0
Anda TB IgG	2	1	1	0
Anda TB IgM	2	1	1	0
ICT	2	1	1	0
In-house	1	1	0	0
Antigen tests	**1**	**1**	**0**	**0**
MTB	1	0	0	0
Cytokine tests	**13**	**10**	**3**	**0**
Interferon-γ	5	5	0	0
Interleukin	4	2	2	0
TNF	3	2	1	0
Total protein	1	1	0	0
Lysozyme tests	**11**	**4**	**1**	**6 p/s**
Lysozyme	10	3	1	6
Lysozyme/total protein ratio	1	1	0	0
Other miscellaneous tests	**6**	**4**	**2**	**0**
CA125	1	1	0	0
Interferon-γ response (PPD)	2	1	1	0
TBSA	1	1	0	0
Lymphocytes	2	1	1	0

PF, pleural fluid; p/s, pleural/serum.

TABLE 30 *Diagnostic test evaluations in pleural TB – summary of key study characteristics for main tests*

	NAAT	Adenosine deaminase	Cytokine	Lysozyme
Total no. of studies	20	42	13	11
Mean sample size	81 (SD 83; range 9–375)	166 (SD 104; range 39–405)	100 (SD 29; range 39–145)	182 (SD 80; range 92–276)
Commercial	6			
Mean sample size	82 (SD 144; range 9–375)			
In-house	14	42	13	11
Mean sample size	80 (SD 46; range 14–140)			
Reference standard				
Culture + clinical diagnosis ± other	11 (55%)	18 (43%)	2 (15%)	9 (82%)
Culture + anti-TB therapy	1 (5%)	5 (12%)	2 (15%)	0
Culture + other	0	15 (36%)	9 (69%)	2 (18%)
Culture alone	7 (35%)	0	0	0
No culture	1 (5%)	4 (10%)	0	0
Disease prevalence (mean, SD, range)	32% (SD 19; range 2–68%)	36% (SD 19%; range 2–68%)	39% (SD 19; range 19–95%)	24% (SD 10; range 7–44%)
Setting				
Hospital-based	15 (75%)	42 (100%)	13 (100%)	11 (100%)
Clinic	0	0	0	0
Laboratory-based	4 (20%)	0	0	0
Unknown	1 (5%)	0	0	0
Sample type				
Pleural fluid	20 (100%)	36 (86%)	9 (69%)	4 (36%)
Serum	0	2 (5%)	4 (31%)	1 (9%)
P/s	0	4 (10%)	0	6 (55%)
Patients representative?				
Yes	8 (40%)	0	0	4 (36%)
No	5 (25%)	21 (50%)	7 (54%)	1 (9%)
Unclear	7 (35%)	21 (50%)	6 (46%)	6 (55%)
Study design prospective?				
Yes	7 (35%)	10 (24%)	3 (23%)	0
No	1 (5%)	2 (5%)	0	7 (64%)
Unclear	12 (60%)	30 (71%)	10 (77%)	4 (36%)
Index test interpreted blinded?				
Yes	7 (35%)	8 (19%)	1 (8%)	1 (9%)
No	0	2 (5%)	0	0
Unclear	13 (65%)	32 (76%)	12 (92%)	10 (91%)
Reference test interpreted blinded?				
Yes	4 (20%)	5 (12%)	0	3 (27%)
No	2	2 (5%)	0	0
Unclear	14 (70%)	35 (83%)	13 (100%)	8 (73%)

evaluations were carried out in 17 separate studies. A summary of key characteristics across all data sets is given in *Table 30*, with details per study provided in Appendix 13 grouped according to evaluation of commercial or in-house tests.

Tests

Most (70%) of the NAAT studies evaluated used in-house tests rather than commercial tests and predominantly used the IS6110 target sequence (9/14). Of the commercial evaluations, three datasets related to the AMTD test, two to Amplicor and one evaluated Amplicis Myco B.

Reference standards

A combined reference standard of culture plus clinical diagnosis (with or without additional tests such as assessment of response to anti-TB therapy) was used in 55% (11/20) of studies (*Table 30*). Culture alone was the reference standard in a further 35% (7/20) and culture combined with a therapy trial in another. The remaining study did not include culture as part of the reference standard but used histology, therapy trial and X-ray.

Sample details

About 75% of studies were hospital based, four (20%) were laboratories reviewing submitted

samples and in one study the setting was not clear. All samples tested in these studies were pleural fluid samples.

Overall, the mean number of patients recruited was 81 (SD 83; range 9–375), and there was little difference between the numbers of patients evaluated in in-house and commercial studies, although the range was wider for the commercial studies (*Table 30*). The mean prevalence of TB was 32% (SD 19%).

We judged the patient sample to have been representative (i.e. the study at least stated that patients 'suspected' of having TB were recruited) in 40% (8/20) of studies. The sample was judged to be unrepresentative in five studies and was unclear in the remaining 12.

Test interpretation

Index test interpretation was reported as blinded in 35% (7/20) of studies and was not clearly reported in the rest. Seven studies were also clearly prospective in design and one was retrospective. Assuming (based on time taken to perform PCR compared with culture) that the index test was interpreted first in the prospective studies and was therefore to all intents and purposes 'blinded' increases the number of studies in which the index test was interpreted blinded to 12 (60%). The reference standard was clearly reported as interpreted blinded in four studies, all of which also reported blinded index test interpretation.

Results

Sensitivity and specificity

The plot of sensitivity against specificity for each study is presented in *Figure 10(a)*. Both sensitivity and specificity are highly variable, especially sensitivity, which ranges from just under 20 to 100%. The tests for heterogeneity for both indices were statistically significant at $p < 0.01$ (*Table 31*). The points remain heterogeneous when split according to commercial or in-house tests (*Figure 11*); although the test for heterogeneity of sensitivities for the commercial tests is non-significant ($p = 0.15$), this is probably due to the low power of the test with such a small number of studies.

Heterogeneity investigation

The heterogeneity investigations (*Table 32*) indicate that only two variables tested had a significant impact on test accuracy. Accuracy was almost 10 times greater in laboratory-based than hospital-based studies (RDOR 9.96; 95% CI: 1.24 to 80.06), and was significantly lower in studies reporting index test blinding (RDOR 0.16; 95% CI: 0.03 to 0.91). In general, once again, accuracy appeared higher in lower prevalence studies, although the impact was not significant.

SROC analysis

SROC curves are shown in *Figure 12*. The overall pooled analysis suggests that the NAAT tests have reasonably high specificity but generally poor sensitivity, regardless of test (*Table 33*). Individually, the AMTD test appears to perform the best, but once two of the five quality criteria are applied, its sensitivity drops from 76 to 33%. Of the hospital-based studies using a combined reference standard (two quality criteria), the in-house tests using the IS6110 target sequence were far superior to any of the commercial tests, although sensitivity was still only 63%. None of the 20 studies met all five study design-related criteria.

Summary

Overall, the studies in this section were not well reported, making it impossible to judge whether certain quality criteria had been met. Study design and blinding of test interpretation were particularly poorly interpreted. This, along with the small number of studies included, may have made it difficult to detect any impact on test accuracy of the various study design features. Nevertheless, those studies that did meet at least two quality criteria demonstrate that the specificity of the NAAT tests with pleural fluid is relatively high (89%), especially when the IS6110 target sequence is used, but sensitivity was almost uniformly poor, indicating that these tests cannot be used reliably to rule out TB infection.

Adenosine deaminase tests

Description and quality of included studies

We included 42 data sets from 36 published studies relating to evaluation of ADA tests. A summary of key characteristics across all data sets is given in *Table 30*, with details per study provided in Appendix 14.

Tests

The most commonly used test was ADA (in 89% of evaluations). Tests used in the remaining evaluations (each test used in one study) included ADA2, ADA/ADA1 ratio, ADA/lysozyme ratio, ADA/total protein and 2′-deoxyadenosine deaminase.

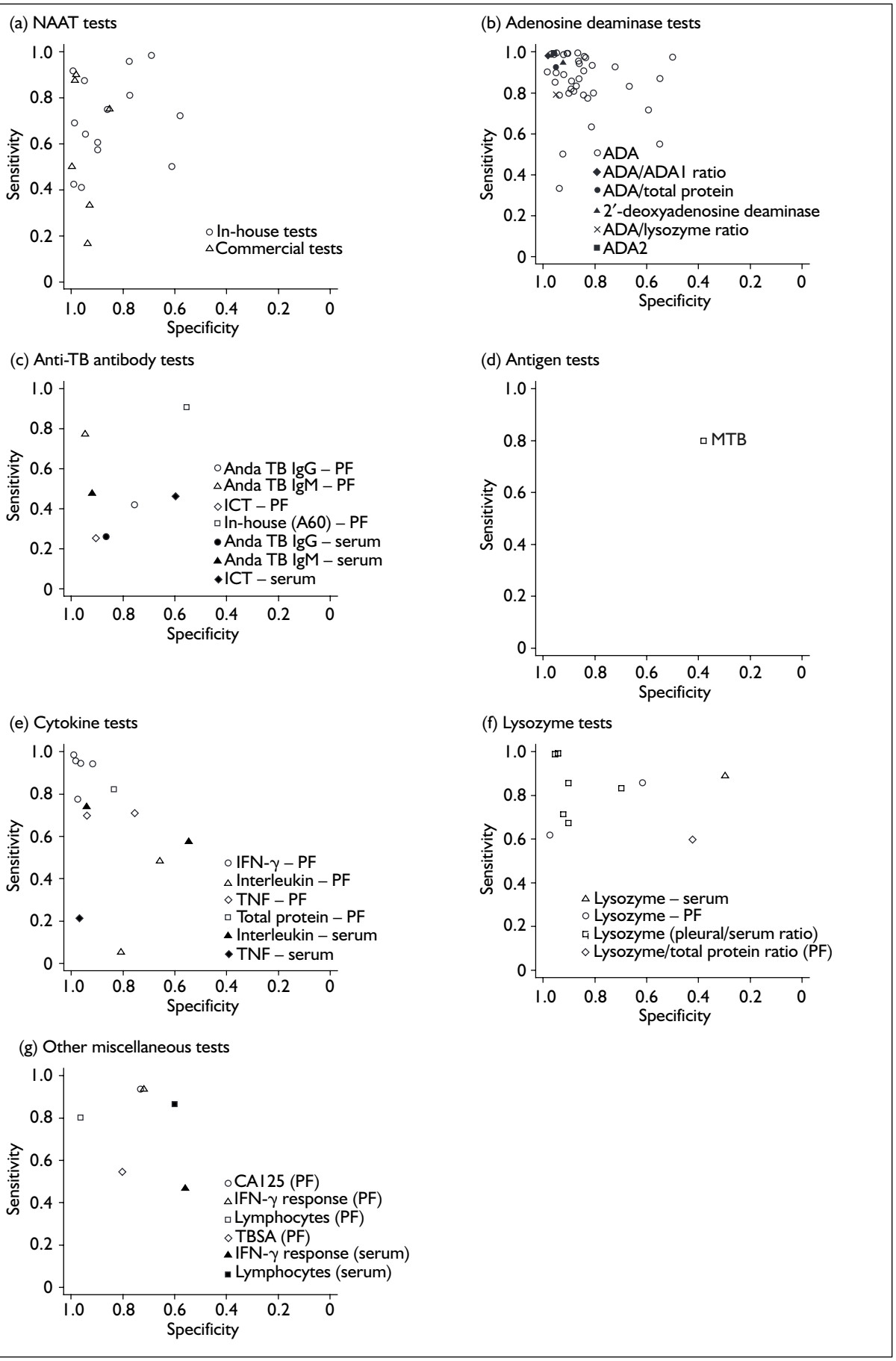


FIGURE 10 *Pleural TB – ROC plots for each group of tests*

TABLE 31 *NAAT tests in pleural TB – tests for heterogeneity of sensitivities and specificities*

		Sensitivity		Specificity	
Comparison	***n***	***Q* statistic**	***p*-Value**	***Q* statistic**	***p*-Value**
All NAAT comparisons	20	56	<0.01	103	<0.01
Commercial tests	**6**	**8**	**0.15**	**13**	**0.02**
AMTD	3	4	0.14	1	0.57
Amplicor	2	1	0.35	3	0.07
Amplicis (Myco B)	1	NA	NA	NA	NA
In-house tests	**14**	**48**	**<0.01**	**72**	**<0.01**
IS6110	9	30	<0.01	23	<0.01
Other targets	5	12	0.02	9	0.06

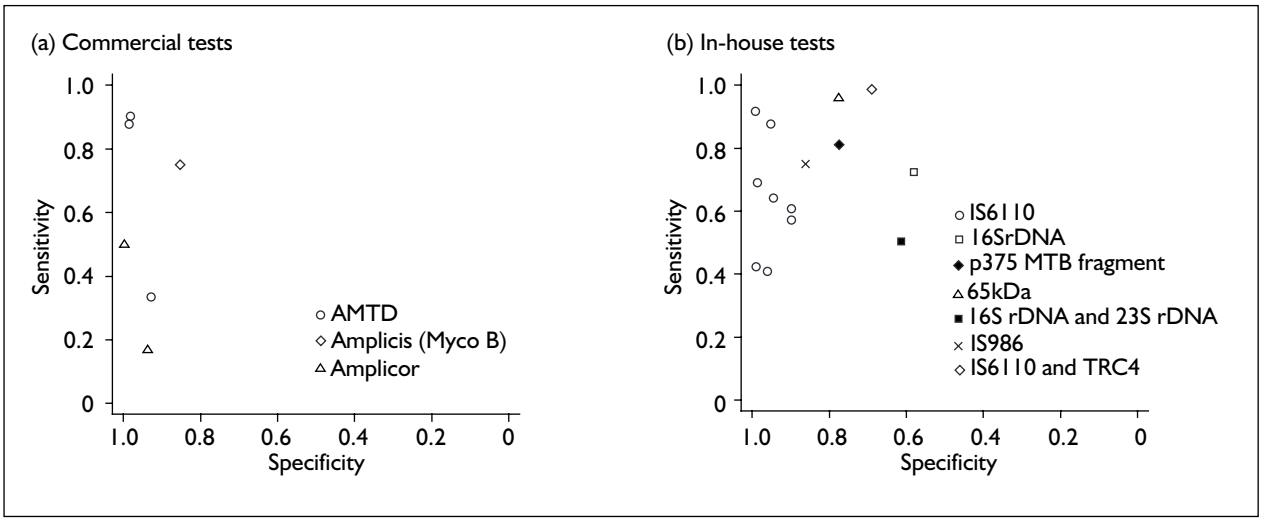


FIGURE 11 *NAAT evaluations in pleural TB – ROC plots by test*

Reference standards

A combined reference standard of culture plus clinical diagnosis (with or without additional tests such as assessment of response to anti-TB therapy) was used in less than half (17/42) of studies; culture alone as a reference standard was not used in any of the studies in this group. Twenty studies used culture plus a trial of anti-TB therapy (n = 5) or culture plus histology (n = 15); in four studies culture was not included in the reference standard (*Table 30*).

Sample details

All of the ADA studies were hospital based, but none were judged to have recruited a representative patient spectrum. This is because of our requirement that studies recruited patients 'suspected' of having TB; most of the studies in this group recruited patients with pleural effusion (i.e. TB was one of the differential diagnoses but not necessarily the main one). Thirty-six studies evaluated pleural fluid samples, two used serum samples and four used both pleural fluid and serum (*Table 30*).

Overall, the mean number of patients recruited was 166 (SD 104; range 39–405) and mean prevalence of TB was 36% (SD 19%).

Test interpretation

Index test interpretation was reported as blinded in 19% (8/42) of studies, was not blinded in two studies and was not clearly reported in the remainder. The reference standard was clearly reported as interpreted blinded in five studies (14%), all of which also reported blinded index test interpretation. Ten studies were clearly prospective in design and two were retrospective.

Results

Sensitivity and specificity

The plot of sensitivity against specificity for each study is presented in *Figure 10(b)*. The sensitivity and specificity pairs are predominantly clustered in the top left-hand quadrant of the graph, with a few outlying studies in terms of both sensitivity and specificity. The tests for heterogeneity for both indices were statistically significant at $p < 0.01$ (*Table 34*). The heterogeneity remains

TABLE 32 *NAAT tests in pleural TB – regression analyses to identify source(s) of heterogeneity*

Comparison	*n* (20)	Model parameters (95% CI) and *p*-value		
		Coefficient	*p*-Value	RDOR[a]
Test type				
AMTD vs in-house	3/14	1.02	0.44	2.78 (0.18 to 42.69)
Amplicor vs in-house	2/14	–0.46	0.80	0.63 (0.01 to 26.89)
Other commercial vs in-house	1/14	–0.69	0.74	0.50 (0.01 to 38.28)
Institute				
Laboratory-based vs hospital-based	4/16	2.30	0.03	9.96 (1.24 to 80.06)
Reference standard				
Culture plus clinical vs culture alone	11/7	–1.18	0.22	0.31 (0.04 to 2.21)
Culture plus other vs culture alone	1/7	1.55	0.53	4.72 (0.03 to 786.00)
No culture vs culture alone	1/7	0.67	0.75	1.95 (0.02 to 155.05)
Design				
Prospective vs retrospective/unknown	7/13	0.22	0.82	1.24 (0.16 to 9.37)
Quality factors				
Patients representative vs unrepresentative/ unknown	7/13	–0.32	0.73	0.73 (0.11 to 4.95)
Index test blinded vs not blinded/unknown	7/13	–1.84	0.04	0.16 (0.03 to 0.91)
Reference test blinded vs not blinded/unknown	4/16	–0.26	0.82	0.77 (0.07 to 7.95)
TB prevalence				
Prevalence <10% vs prevalence 30+%	4/11	0.98	0.42	2.67 (0.21 to 33.26)
Prevalence 10–20% vs prevalence 30+%	1/11	–1.67	0.44	0.19 (0.00 to 16.97)
Prevalence 20–30% vs prevalence 30+%	4/11	0.75	0.54	2.11 (0.17 to 26.38)
% patients smear positive				
<5% smear positive vs >30%	6/4	–0.56	0.70	0.57 (0.02 to 14.62)
5–10% smear positive vs >30%	3/4	0.24	0.89	1.27 (0.03 to 53.59)

[a] Relative diagnostic odds ratio, i.e. in one group compared with the other.

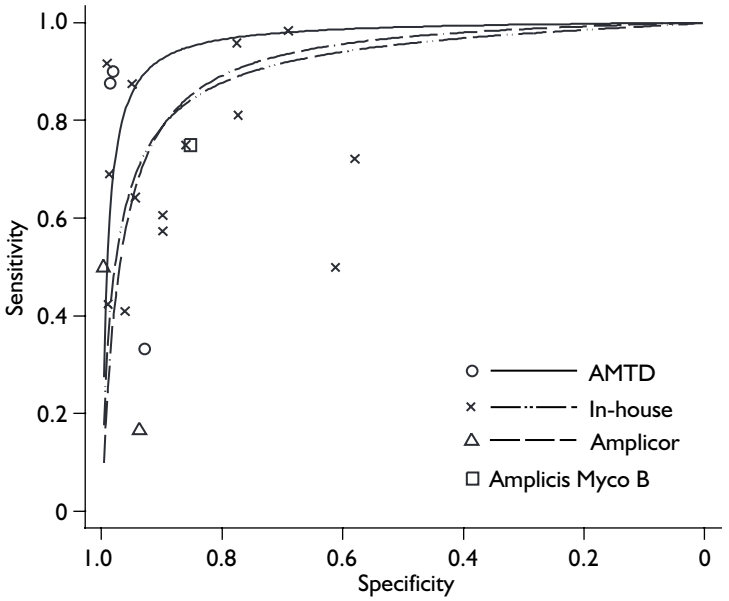
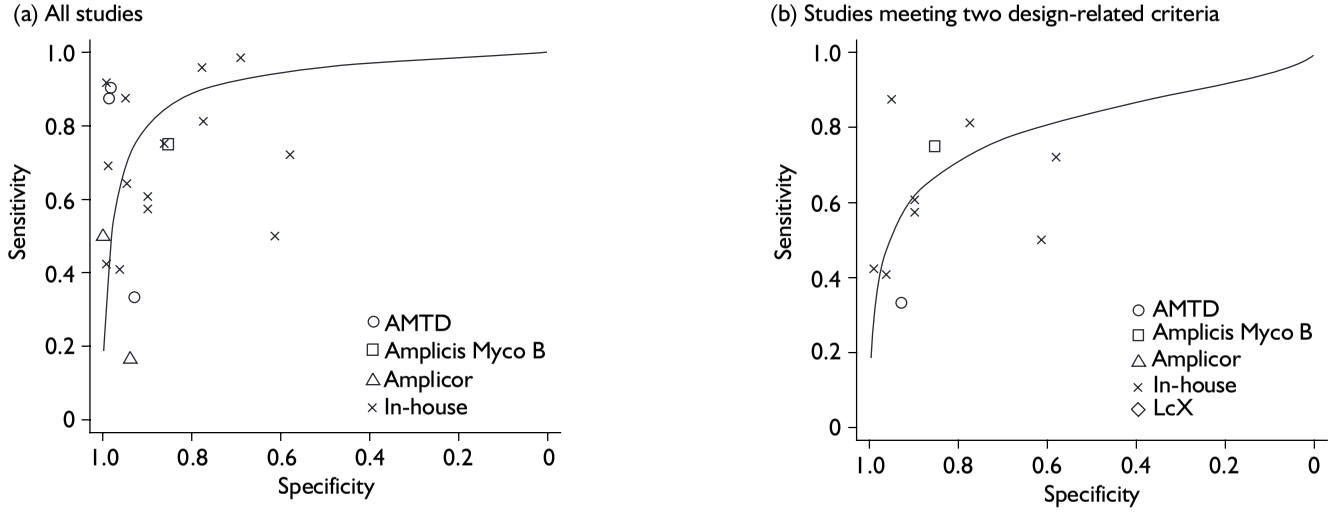


FIGURE 12 *NAAT evaluations in pleural TB – SROC curves by test*

TABLE 33 *NAAT tests in pleural TB – SROC regression models and associated SROC curves*



	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se^d (%)	Sp^d (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se^d (%)	Sp^d (%)
All tests	20	33.18 (10.57 to 104.14)	0.61	72.6	93.7	10	8.50 (2.35 to 30.70)	0.25	61.8	89.5
All commercial	**6**	**59.33 (0.15 to 23777.38)**	**0.97**	**62.0**	**97.1**	**2**	**10.63 (0.02 to 5540.47)**	**NA**	**55.1**	**89.7**
AMTD	3	113.55 (0.24 to 53465	NA	76.0	97.3	1	6.5 (0.28 to 151.13)	NA	33.3	92.9
Amplicor	2	33.14 (0.00 to >10,000)	NA	30.9	98.7	0				
Amplicis Myco B	1	17.40 (0.54 to 559.58)	NA	75.0	85.3	1	17.40 (0.54 to 559.58)	NA	75.0	85.3
In-house	**14**	**30.47 (9.35 to 99.28)**	**0.30**	**76.5**	**91.3**	**8**	**7.89 (1.69 to 36.81)**	**0.19**	**63.5**	**89.5**
IS6110	9	78.17 (14.01 to 436.07)	0.75	75.0	95.7	5	27.77 (0.65 to 1187.75)	0.93	59.9	95.4
Other target sequences	5	7.98 (0.94 to 67.69)	0.30	79.1	73.4	3	4.38 (0.27 to 72.27)	NA	69.1	66.2
Subgroups meeting each design-related criterion										
Hospital-based	16	27.75 (9.36 to 82.29)	0.72	70.6	91.2					
Combined reference test	11	10.75 (1.79 to 64.61)	0.28	65.9	91.5					
Index test blinded	12	8.25 (1.72 to 39.67)	0.12	64.5	92.1					
Reference test blinded	4	30.47 (4.58 to 202.64)	NA	75.3	90.9					
Patients representative	8	10.68 (0.43 to 265.01)	0.27	66.2	94.6					

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

TABLE 34 *Adenosine deaminase evaluations in pleural TB – tests for heterogeneity of sensitivities and specificities*

		Sensitivity		Specificity	
Comparison	***n***	***Q* statistic**	***p*-Value**	***Q* statistic**	***p*-Value**
All adenosine deaminase data sets	42	200	<0.01	358	<0.01
By test					
ADA	37	182	<0.01	278	<0.01
ADA2	1	NA		NA	
ADA/ADA1	1	NA		NA	
ADA/lysozyme	1	NA		NA	
ADA/total protein	1	NA		NA	
2′-Deoxyadenosine deaminase	1	NA		NA	

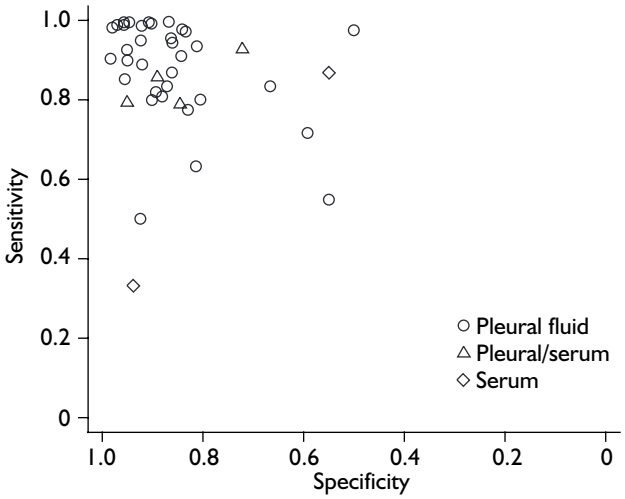


FIGURE 13 *Adenosine deaminase evaluations in pleural TB – ROC plot by sample*

TABLE 35 *ADA tests in pleural TB – regression analyses to identify source(s) of heterogeneity*

Comparison	***n* (36)**	**Model parameters (95% CI) and *p*-value**		
		Coefficient	***p*-Value**	**RDOR[a]**
Sample type				
Pleural/serum vs pleural fluid	3/32	–0.68	0.61	0.51 (0.04 to 7.25)
Pleural/serum vs pleural fluid	1/32	–1.69	0.36	0.18 (0.00 to 7.42)
Serum vs pleural fluid	3/32	–1.79	0.19	0.17 (0.01 to 2.47)
Institute				
Hospital setting vs non-hospital	36/0			
Reference standard				
Culture plus other vs culture plus clinical	19/16	0.38	0.54	1.46 (0.42 to 5.11)
No culture vs culture plus clinical	2/16	–0.96	0.48	0.38 (0.02 to 5.97)
Design				
Prospective vs retrospective/unknown	9/27	–1.27	0.06	0.28 (0.07 to 4.06)
Quality factors				
Patients *not* representative vs unknown	19/18	1.16	0.05	3.20 (1.01 to 10.20)
Index test blinded vs not blinded/unknown	8/29	–0.59	0.42	0.56 (0.13 to 2.36)
Reference test blinded vs not blinded/unknown	5/31	–0.94	0.30	0.40 (0.07 to 2.32)
TB prevalence				
Prevalence <10% vs prevalence 30+%	1/22	–0.31	0.87	0.74 (0.02 to 34.67)
Prevalence 10–20% vs prevalence 30+%	2/22	0.11	0.94	1.11 (0.07 to 17.87)
Prevalence 20–30% vs prevalence 30+%	12/22	0.33	0.63	1.39 (0.36 to 5.34)

[a] Relative diagnostic odds ratio, i.e. in one group compared with the other.

when the tests are restricted to ADA only. The same ROC plot is presented in *Figure 13* according to type of sample; test performance appears best in pleural fluid samples and worst in serum samples.

Heterogeneity investigation

The heterogeneity investigations restricted to evaluations of the ADA test (*Table 35*) indicate that the only variable to approach statistical significance was whether or not the sample was judged to be representative. Samples that were clearly unrepresentative had three times higher accuracy than those in which sample representativeness was judged 'unclear' (RDOR 3.20; 95% CI: 1.01 to 10.20, $p = 0.04$). Studies that were prospective in design had DORs one-fifth of those of retrospective or unknown designs (RDOR 0.28; 95% CI: 0.07 to 4.06, $p = 0.06$).

SROC analysis

The overall pooled analysis suggests that the ADA tests have reasonably high sensitivity and specificity, although specificity drops slightly when studies are reduced by applying two of the quality criteria (*Table 36*). Looking at only the ADA studies, summary sensitivity and specificity when two design-related criteria are applied are 92.4 and 86.5%, respectively. None of the studies met all five design-related criteria. Specificity appears higher in the other tests in the ADA group, but as only one data set per test is available no strong conclusions can be drawn from this.

Table 37 shows a restricted analysis for the ADA test in different clinical samples. ADA performs better in pleural fluid than in any of the other samples or sample combinations. Results for subgroups according to each of the five selected study design-related criteria are given in *Table 38*. The DOR was higher in studies using a combined reference standard but lower in those reporting blinded test interpretation.

Summary

Overall, the studies in this section were not well reported, especially in terms of test interpretation blinding and study design, making an accurate evaluation of how good the tests are difficult. Lack of blinding of test interpretation in particular can lead to overestimation of accuracy. However, all were hospital based and those using a combined reference standard demonstrated reasonably good accuracy. The impact of using unrepresentative samples perhaps was surprising given that none of the other regression analyses have found studies with representative samples to be significantly less accurate than other studies. However, the pleural TB studies were particularly likely to recruit patients with a particular clinical feature (in this case pleural effusion) rather than patients with a given set of signs and symptoms suggestive of TB infection. The data for the subgroup of ADA studies with unrepresentative samples indicates that most of the impact from this factor is on specificities, that is, it is the make-up of the control groups of the studies that differs. Although we have no representative studies with which to make a comparison, it may be that the ADA test is less accurate when applied to groups of patients specifically suspected of pleural TB infection.

Other serodiagnostic and biochemical tests

Description and quality of included studies

We included 35 data sets relating to other tests that have been used to detect the presence of TB infection (*Table 29*). A summary of study characteristics for the cytokine and lysozyme tests are provided in *Table 30* with a further breakdown of the cytokine tests in *Table 39*. Summary details per study are provided in Appendix 15.

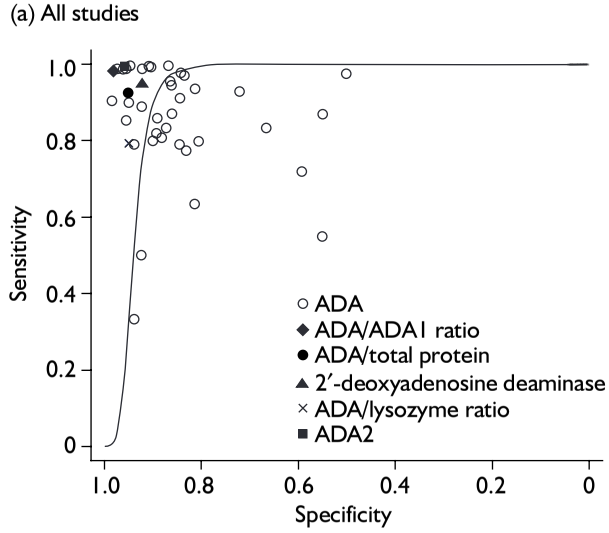
Anti-TB antibody tests

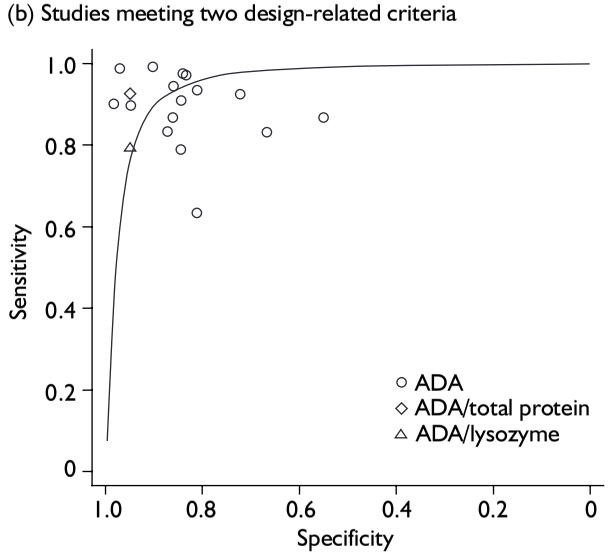
The seven anti-TB antibody test evaluations were carried out in three studies, two evaluating two commercial tests in two different samples and one of an in-house test to detect antibodies to the A60 antigen (Appendix 15). Sample sizes in the three studies were between 81[206] and 125[207] and TB prevalence ranged from 56%[208] to 70%.[207] The commercial tests (ICT,[208] Anda TB IgG[207] and Anda TB IgM[207]) were each evaluated in serum and in pleural fluid samples. The in-house test evaluation[206] was carried out with pleural fluid samples only. The ICT study was laboratory based and the Anda TB and in-house test studies were hospital based (although patients were not judged to be representative in the latter). All three studies used culture plus histology as the reference standard; the in-house test study also included an anti-TB therapy trial. Study design and test interpretation were not clearly reported apart from index test blinding in the ICT study.[208]

Antigen tests

One study of an antigen test (using *M. TB*) was identified.[209] The study was small with only 36 patients, but was otherwise of high quality. A combined reference standard was used; it was

TABLE 36 *Adenosine deaminase evaluations in pleural TB – SROC models and associated SROC curves*

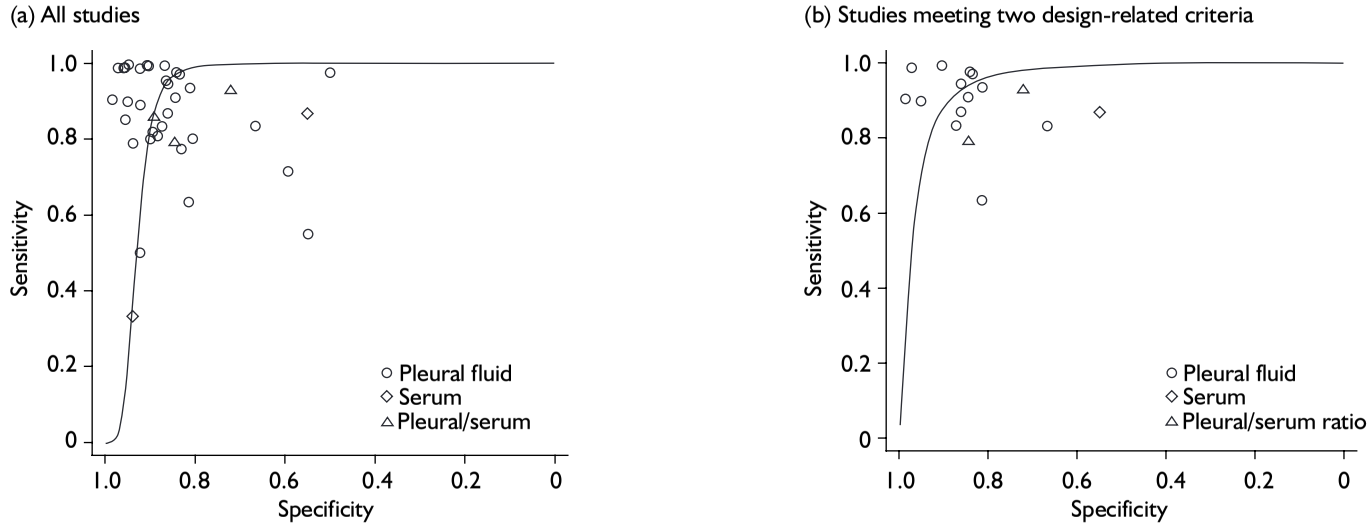




Comparison	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
All data sets	42	78.54 (43.45 to 141.96)	<0.01	92.4	89.1	18	77.32 (31.91 to 187.33)	0.68	91.9	87.9
By test										
ADA	36	61.25 (33.12 to 113.28)	<0.01	91.8	87.7	16	68.33 (23.60 to 197.83)	0.59	92.4	86.5
ADA2	1	3505.70 (204.23 to 60176.61)	NA	99.4	95.8	0				
ADA/ADA1	1	2768.33 (109.48 to 7003.56)	NA	98.2	98.1	1	2768.33 (109.48 to 7003.56)	NA	98.2	98.1
ADA/lysozyme	1	71.85 (29.24 to 176.58)	NA	79.2	95.0	1	71.85 (29.24 to 176.58)	NA	79.2	95.0
ADA/total protein	1	230.52 (70.41 to 754.72)	NA	92.4	95.0	0				
2′-Deoxyadenosine deaminase	1	216.86 (72.12 to 652.03)	NA	94.7	92.3	0				

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

TABLE 37 *ADA test in pleural TB – SROC models by sample type*



	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se^d (%)	Sp^d (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se^d(%)	Sp^d(%)
All tests	36	61.25 (33.12 to 113.28)	<0.01	91.8	87.7	16	68.33 (23.60 to 197.83)	0.59	92.4	86.5
Pleural fluid	32	70.85 (35.59 to 141.02)	0.01	92.9	88.4	13	90.24 (28.63 to 284.44)	0.39	93.2	88.8
Pleural/serum	2	31.49 (0.13 to 7570.88)	NA	82.6	86.9	2	99.60 (30.17 to 328.75)	NA	87.3	79.0
Serum	2	7.77 (4.97 to 12.15)	NA	64.4	81.1	1	8.05 (3.48 to 18.62)	NA	86.8	55.0

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

TABLE 38 *ADA test in pleural TB – effect of individual quality criteria on test accuracy*

	n[a]	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
All tests	36	61.25 (33.12 to 113.28)	<0.01	91.8	87.7
Hospital-based	36	61.25 (33.12 to 113.28)	<0.01	91.8	87.7
Culture + clinical reference test	16	68.33 (23.60 to 197.873)	0.59	92.4	86.5
Index test blinded/study prospective	11	39.83 (10.32 to 153.82)	0.42	89.6	84.3
Reference test blinded	5	26.88 (1.13 to 638.72)	0.62	81.0	85.0
Patients representative	0				
Patients *not* representative	19	102.59 (47.50 to 221.9)	<0.01	92.0	90.9

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

TABLE 39 *Cytokine evaluations in pleural TB – summary of key study characteristics*

	Interferon-γ	Interleukin	TNF
Tests (*n*; mean sample size, SD, range)	5 93.4 (SD 46; range 39–145)	4 97 (SD 4; range 93–102)	3 107 (SD 17; range 97–127)
Reference standard			
Culture + clinical diagnosis ± other	1 (20%)	0	1 (33%)
Culture + anti-TB therapy	1 (20%)	0	0
Culture + other	3 (60%)	4 (100%)	2 (67%)
Culture alone	0	0	0
Clinical diagnosis alone	0	0	0
Disease prevalence (mean, SD, range)	38% (SD 14; range 24–59%)	29% (SD 7; range 19–34%)	33% (SD 1; range 32–34%)
Setting			
Hospital-based	5 (100%)	4 (100%)	3 (100%)
Laboratory-based	0	0	0
Unknown	0	0	0
Sample type			
Pleural fluid	5 (100%)	2 (50%)	2 (67%)
Serum	0	2 (50%)	1 (33%)
Patients representative?			
Yes	0	0	0
No	2 (40%)	3 (75%)	2 (67%)
Unclear	3 (60%)	1 (25%)	1 (33%)
Study design prospective?			
Yes	2 (40%)	0	0
No	0	0	0
Unclear	3 (60%)	4 (100%)	3 (100%)
Index test interpreted blinded?			
Yes	1 (20%)	0	0
No	0	0	0
Unclear	4 (80%)	4 (100%)	3 (100%)
Reference test interpreted blinded?			
Yes	0	0	0
No	0	0	0
Unclear	5 (100%)	4 (100%)	3 (100%)

prospective in design and blinded index and reference test interpretation was reported (Appendix 15). It was not possible to judge whether the patients were representative or not.

Cytokine tests

Twelve cytokine studies were identified, including five on interferon-γ, four on interleukin and three evaluating TNF (*Table 39*).

The five interferon-γ evaluations each came from separate studies (Appendix 15). All were carried out on pleural fluid and all were hospital based. The mean sample size was 93 (SD 46) and TB prevalence was 38% (SD 14%) (*Table 39*). A combined reference standard of culture plus clinical diagnosis and other diagnostic tests was used in only one study; the others used culture and histology, one also including anti-TB treatment outcome. Two of the studies were prospective in design, one used blinded index test interpretation and none reported blinded reference test interpretation. Patient samples were judged to be clearly unrepresentative in three studies.

The four interleukin evaluations were carried out in three separate studies (Appendix 15). The specific tests evaluated were IL-1 in pleural fluid and in serum (97 patients), IL-1B in pleural fluid (102 patients) and IL-2R in serum (93 patients). The mean sample size was 97 (SD 4) and TB prevalence was 29% (SD 7%) (*Table 39*). All studies used culture plus histology as the reference standard, all were hospital based and none reported blinded test interpretation or study design information. The patient samples were judged to be unrepresentative in all three studies.

The three TNF evaluations were carried out in two separate studies, one using pleural fluid samples only and the other reporting results for both pleural fluid and serum samples (Appendix 15). The mean sample size was 107 (SD 17) and TB prevalence was 33% (SD 1%) (*Table 39*). One study used culture plus histology as the reference standard whereas the other used a combined reference standard that included culture and clinical diagnosis. Neither reported blinded test interpretation or study design information. The patient samples were judged to be unrepresentative in one study and not clear in the other.

Lysozyme tests

Eleven lysozyme evaluations, carried out in seven studies, were identified (Appendix 15).

Four evaluations were carried out using pleural fluid (*Table 30*). Mean sample size was 193 (SD 92) and TB prevalence was 24% (SD 8%). All studies were hospital based and three used a combined reference standard of culture plus clinical diagnosis and other diagnostic tests. One study used blinded reference test interpretation, but otherwise design-related criteria were not well reported. Patient samples were judged to be clearly unrepresentative in three of the four studies.

Six evaluations were carried out in both pleural fluid and serum (pleural/serum lysozyme ratio) (*Table 30*). The mean sample size was 160 (SD 74%) and TB prevalence was 25% (SD 13%). Five of the six studies used a combined reference standard and the remaining one used culture plus histology. All were hospital based; one was prospective and reported both index and reference test blinding. One further study also reported blinded reference test interpretation. The patient samples were judged to be clearly unrepresentative in four studies.

The single study evaluating lysozyme in serum alone used a combined reference standard and was hospital based. The patient sample was judged to be unrepresentative but none of the other study design features could be determined from the study report.

Miscellaneous other tests

The studies evaluating miscellaneous other tests (Appendix 15) were generally small and poorly reported. Sample sizes in the four studies ranged

TABLE 40 *Serodiagnostic and biochemical test evaluations in pleural TB – tests for heterogeneity of sensitivities and specificities*

		Sensitivity		Specificity	
Comparison	***n***	***Q statistic***	***p*-Value**	***Q statistic***	***p*-Value**
Anti-TB antibody tests	7	81	<0.01	31	<0.01
Antigen tests	1	NA		NA	
Cytokine tests	13	65	<0.01	79	<0.10
Lysozyme tests	11	34	<0.01	336	<0.01
Other miscellaneous tests	6	12	0.04	9	0.11

from 19 to 40 with prevalences of TB from 37 to 52%. All were hospital based, none used a combination of culture plus clinical diagnosis as a reference standard and it was not possible to judge the study design or whether or not blinded test interpretation had been used. In one study, the patient sample was clearly unrepresentative.

Results

The plot of sensitivity against specificity for each group of tests is presented in *Figure 10(c–g)*. *Table 40* provides the results of the heterogeneity tests of sensitivities and specifities in each group of tests.

Anti-TB antibody tests

The anti-TB antibody tests generally performed poorly (*Table 41*) with a summary DOR of 6.72 (95% CI: 1.05 to 43.10). For the most part, sensitivity was poor and specificity slightly higher, although this was reversed for the in-house test, which had higher sensitivity (91%) and lower specificity (56%). None of the studies met the first two design-related criteria that we applied: hospital based and combined reference standard used.

Antigen test

The antigen test evaluated had relatively high sensitivity (80%) but specificity was very poor at only 38.1% (*Table 42*).

Cytokine tests

The best performing cytokine test was interferon-γ (*Table 43*), with a DOR of 817.48 (95% CI: 32.93 to 20,296.20) and associated sensitivity and specificity of 94.5 and 97.3%, respectively. Only one of the five evaluations met two of our design-related criteria and sensitivity was much lower at just 77.6%. None of the studies met all five design-related criteria.

Interleukin and TNF were generally not found to be good indicators of TB infection (*Table 43*). One study[210] of Il-2R in 93 patients did achieve sensitivity of 74% and specificity 94%, but no blinding of test interpretation was reported and the patient sample was judged not to be representative.

Lysozyme tests

Of the lysozyme evaluations, those estimating the pleural/serum lysozyme ratio were the most accurate, producing both sensitivity and specificity estimates of just over 90% (*Table 44*). Five of the six studies in this group were both hospital based and used an appropriate combined reference standard. The summary DOR for these studies was 87.11 (95% CI: 4.51 to 1683.27), with a sensitivity of 93.8% and a specificity of 90.7%. None of the studies met the remaining three design-related criteria. Lysozyme estimation in pleural fluid samples and in serum were not found to be good indicators of TB infection.

Miscellaneous other tests

The remaining tests were each evaluated in only one or two studies, and none appear to offer a great deal of promise for detecting TB infection (*Table 45*). The highest specificity was seen in the study assessing lymphocytes in pleural fluid (96%), with a corresponding sensitivity of 80%,[211] but with only 40 patients, a potentially inadequate reference standard and no blinded test interpretation reported, this is unlikely to be a true representation of its true accuracy.

Summary

Two of the tests covered in this section may potentially have some role in the diagnosis of pleural TB infection: interferon-γ in pleural fluid and the pleural/serum lysozyme ratio. The number of reasonable quality studies in both cases was small, and none reported using index or reference test blinding, which, if not implemented, can inflate test accuracy. Specificity of the interferon-γ test was particularly high at 97.5%.

Empirical comparisons between test types

NAAT test comparisons

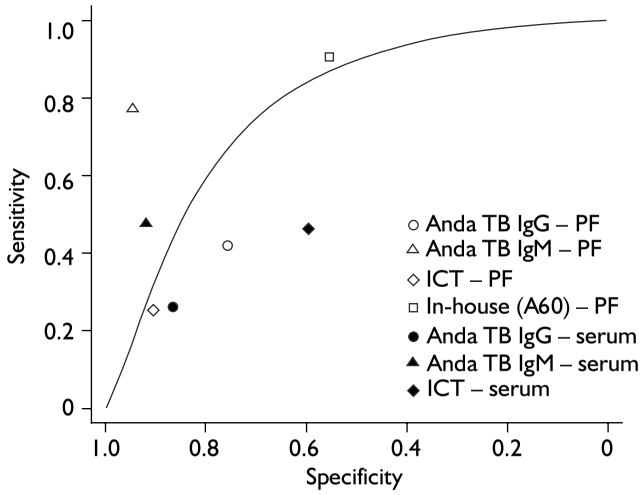
The results of the unpaired comparisons between different types of test (*Table 46*) suggest that the NAAT tests are superior in accuracy to the anti-TB antibody tests, antigen tests and cytokine tests, inferior to adenosine deaminase tests and equivalent to the lysozyme tests, although the result was significant only for the unpaired comparison between NAAT tests and anti-TB antibody tests ($p = 0.02$).

Paired data (more than one test type evaluated in the same study) are available for two of the NAAT test comparisons and in both cases the result conflicts with the unpaired comparison result.

First, the paired data appear to show that the NAAT tests actually perform better than the ADA tests. The two studies[212,213] providing these data both evaluated an in-house NAAT test using the IS6110 target sequence compared with the ADA test to give direct comparisons of four NAAT data sets (one study[213] reported three separate molecular amplification results) with two ADA data sets. The ADA test performed particularly poorly

TABLE 41 *Anti-TB antibody tests in pleural TB – SROC model and SROC curve*

(a) All studies



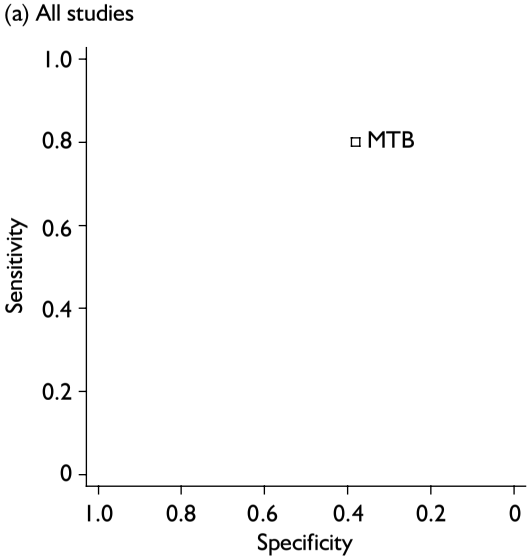
(b) Studies meeting two design-related criteria

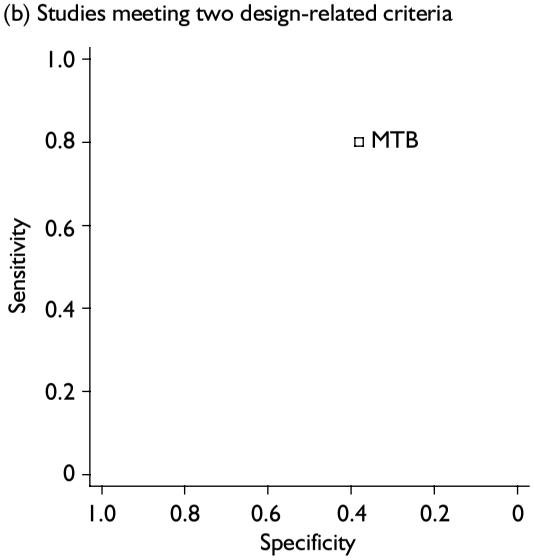
None

Comparison	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
Anti-TB antibody tests	**7**	**6.72 (1.05 to 43.10)**	**0.70**	**53.0**	**83.1**
Anda TB IgG	2	2.26 (2.21 to 2.31)	NA	33.6	81.7
Anda TB IgM	2	24.81 (0.00 to 1663143.43)	NA	63.8	93.4
ICT	2	2.02 (0.01 to 705.05)	NA	35.1	78.8
In-house (A60)	1	12.25 (3.72 to 40.38)	NA	90.7	55.6
All pleural fluid	4	8.52 (0.80 to 90.77)	NA	62.9	83.4
All serum	3	3.10 (0.21 to 45.69)	NA	39.5	82.6

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

TABLE 42 *Antigen tests in pleural TB – SROC model and SROC curve*

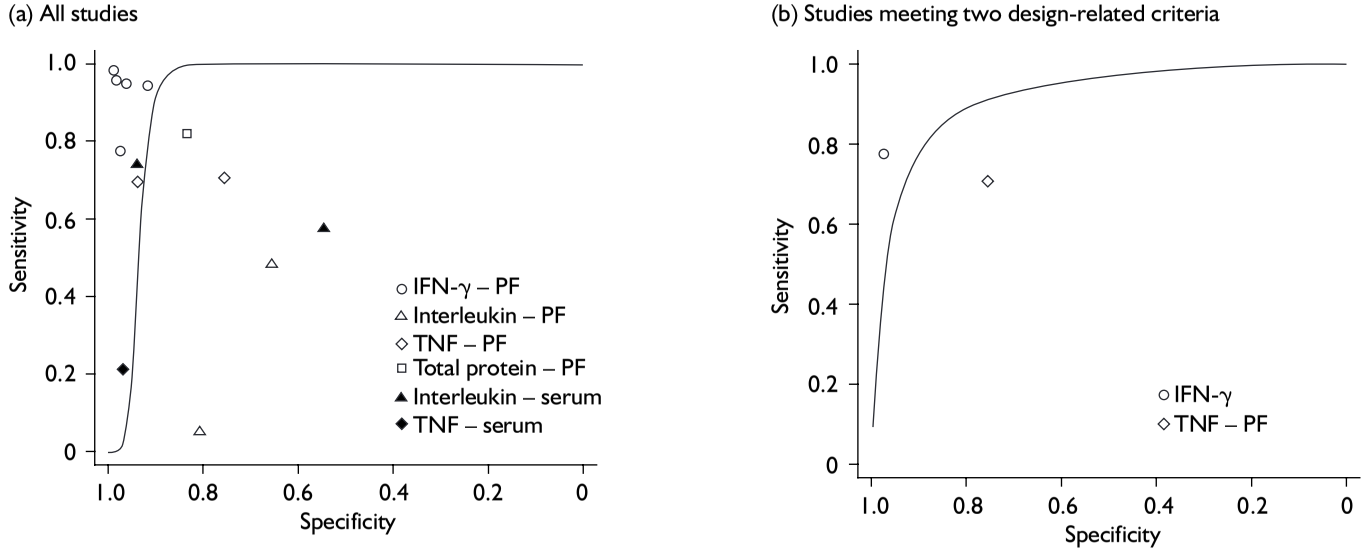




Comparison	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
MTB[e]	1	2.46 (0.53 to 11.50)		80.0	38.1	1	2.46 (0.53 to 11.50)		80.0	38.1

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.
[e] Includes three pleural TB data sets, one peritoneal and two in pulmonary TB.

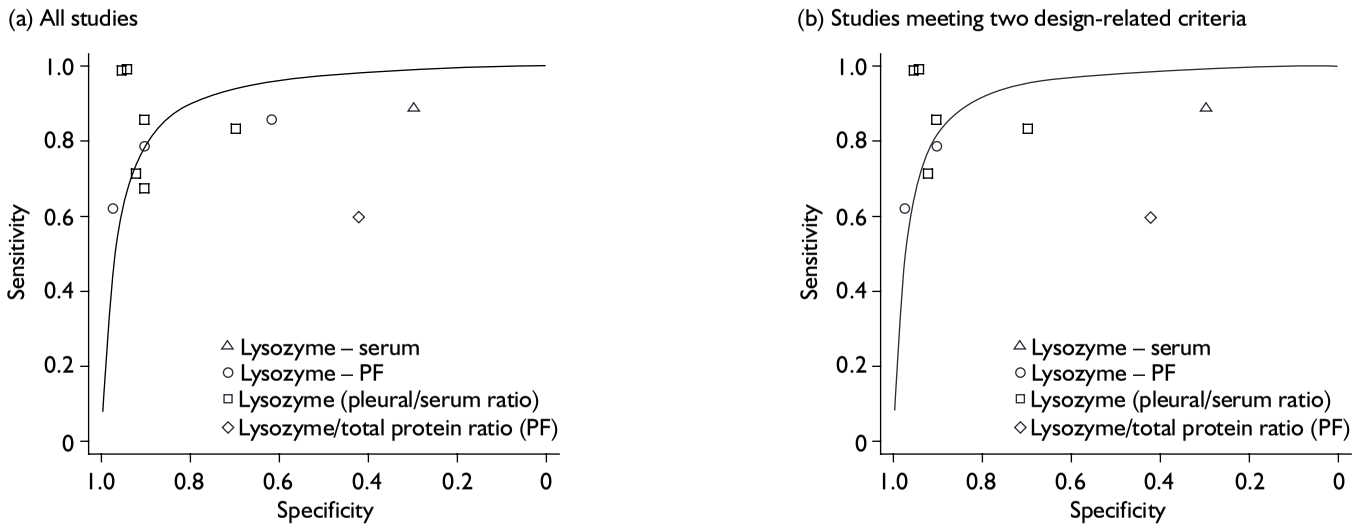
TABLE 43 *Cytokine tests in pleural TB – SROC model and SROC curve*



Comparison	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
All data sets	13	85.31 (9.18 to 792.46)	0.21	75.4	92.1	2	31.57 (0.00 to >2,000,000)	NA	74.3	91.6
By test										
Interferon-γ	5	817.48 (32.93 to 20,296.20)	0.70	94.5	97.3	1	133.27 (28.76 to 617.61)	NA	77.6	97.5
Interleukin	4	2.35 (0.07 to 74.55)	NA	40.2	77.8	0				
TNF	3	12.91 (1.55 to 107.41)	NA	53.4	91.9	1	7.48 (3.25 to 17.21)	NA	70.7	75.6
Total protein		23.00 (2.55 to 207.75)	NA	82.1	83.3					
By sample										
Pleural fluid	10	185.68 (12.83 to 2686.64)	0.14	81.0	92.8	2	As above			
Serum	3	8.46 (0.14 to 508.15)	NA	50.4	89.3	0				

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

TABLE 44 *Lysozyme tests in pleural TB: SROC model and SROC curve*

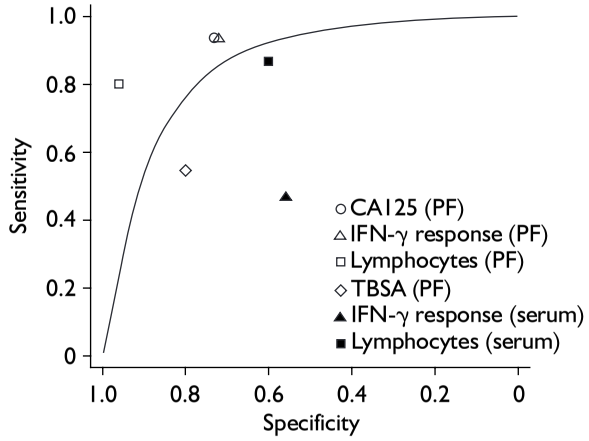


Comparison	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
All data sets	11	34.67 (6.56 to 183.24)	0.94	86.4	84.6	9	42.47 (5.01 to 359.79)	0.92	87.9	85.6
Pleural fluid samples	4	12.07 (0.71 to 205.82)	NA	72.9	81.7	3	13.00 (0.06 to 2890.47)	NA	67.4	86.3
Pleural/serum ratio	6	91.47 (13.12 to 637.83)	0.09	91.5	90.6	5	87.11 (4.51 to 1683.27)	0.18	93.8	90.7
Serum	1	3.33 (1.36 to 8.17)	NA	88.7	70.2	1	3.33 (1.36 to 8.17)	NA	88.7	70.2

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.
[e] Includes three pleural TB data sets, one peritoneal and two in pulmonary TB.

TABLE 45 *Other miscellaneous tests in pleural TB – SROC model and SROC curve*

(a) All studies



(b) Studies meeting two design-related criteria

None

Comparison	*n*[a]	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
Pleural TB[e]	6	13.47 (1.67 to 108.85)	0.80	80.9	76.7
CA125	1	40.71 (1.81 to 916.58)		93.7	73.1
Interferon-γ response (PPD)	2	6.33 (0.00 to >1,000,000)	NA	77.8	64.4
TBSA	1	4.80 (0.68 to 33.80)		54.5	80.0
Lymphocytes	2	30.59 (0.00 to >1,000,000)	NA	83.6	85.7
All tests					
Pleural fluid samples	4	28.67 (3.81 to 783.97)	NA	84.9	83.6
Serum samples	2	3.30 (0.00 to >1,000,000)	NA	70.5	58.0

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.
[e] Includes three pleural TB data sets, one peritoneal and two in pulmonary TB.

TABLE 46 *Pleural TB – comparisons between test types*[a]

		NAAT (*n* = 20)			**Adenosine deaminase (*n* = 42)**			**Anti-TB antibody (*n* = 9)**	
Adenosine deaminase	20/42	ROR = 1.50, *p* = 0.51	←						
	4/2	*ROR = 0.29*, p = *0.43*	↑						
Anti-TB antibody	20/7	ROR = 0.14, *p* = 0.02	↑↑	42/7	ROR=0.14, *p* = 0.02	↑↑			
	0/0	–	–	1/1	–	–			
Antigen	20/1	ROR = 0.09, *p*= 0.26	↑	42/1	ROR = 0.01, *p* = 0.02	↑↑	7/1	ROR = 0.45, *p* = 0.59	↑
	0/0	–	–	0/0	–	–	0/0	–	–
Cytokine	20/13	ROR = 0.87, *p* = 0.87	↑	42/13	ROR = 1.02, *p* = 0.98	↔	7/13	ROR = 6.04, *p* = 0.13	←
	3/1	*ROR = 9.37*, p = *0.01*	←←	*9/10*	*ROR = 0.67*, p = *0.67*	↑	0/0	–	–
Lysozyme	20/11	ROR = 1.00, *p* = 1.00	↔	42/11	ROR = 0.39, *p* = 0.17	↑	7/11	ROR = 5.62, *p* = 0.14	←
	0/0	–	–	*8/7*	*ROR = 0.11*, p = *0.08*	↑		–	–
Individual test comparisons					**ADA**				
			IFN-γ	37/5	ROR = 15.68, *p* < 0.01	←←			
				9/3	*ROR = 8.33*, p = *0.06*	←		**ICT**	
						IFN-γ	2/5	ROR = 189.78, *p* < 0.01	←←
							0/0	–	–
		Antigen			**Cytokine**				
	***n* = 1**	***n* = 13**							
Cytokine	1/13	ROR = 118.53, *p* = 0.18	←						
	0/0	–	–						
Lysozyme	1/11	ROR = 15.03, *p* = 0.33	←	13/11	ROR = 0.59, *p* = 0.66	↑			
	0/0	–	–	*1/2*	*ROR = 0.07*, p = *0.14*	↑			

ROR, relative odds ratio – ratio of DORs.

[a] ↑, ←, arrows point towards test with higher DOR; ↔ indicates tests equivalent. Upper cells indicate unpaired comparisons; lower cells (denoted in italics) indicate results from paired comparisons.

in one of these studies,[212] with sensitivity and specificity both at 55% (Appendix 14). This study was prospective in design and used blinded test interpretation; however, it was also small with only 60 patients and the reference test used was histological findings, smear status or suggestive chest X-ray plus response to treatment. The second study[213] was larger (140 patients), also prospective in design and used a combined reference standard of culture plus histology, clinical diagnosis and response to treatment, although blinding was not described, but sensitivity and specificity were higher at 87 and 86%, respectively. The conflicting results from these two studies make it difficult to draw any definite conclusion regarding the superiority of the ADA test; however, it is notable that once studies are restricted to those that are hospital based and used a combined reference standard, the DOR for the remaining ADA studies is 68.33 (*Table 37*) compared with 27.77 (*Table 33*) for the in-house NAAT tests, suggesting that in fact the ADA tests are superior in performance.

The paired data also appear to show that the cytokine tests actually perform better than the NAAT tests (ROR 9.37, $p = 0.01$) (*Table 46*). These data came from a single study[213] also reported above, which compared three separate molecular amplification results with the interferon-γ test. The interferon-γ test was superior to all of the NAAT test evaluations in terms of both sensitivity and specificity. On the basis of a single study it is difficult to draw any stronger conclusion regarding test superiority; however, if we look at the results for the five interferon-γ test evaluations separately from the cytokine group as a whole, the individual SROC DOR is 817.48 (*Table 43*) compared with 78.17 for the nine in-house IS6110 NAAT evaluations (*Table 33*), suggesting that the former test is in fact superior in accuracy.

Finally, the NAAT tests appear to perform equivalently to the group of lysozyme tests (ROR > 0.99), with individual summary DORs of 33.18 and 34.67, respectively (*Tables 33* and *44*). The estimated summary sensitivities and specificities from these DORs, however, are very different: for the NAAT tests they are 72.6 and 93.7% and for the lysozyme tests 86.4 and 94.6%. Although the tests appear equivalent, the most appropriate test may depend on whether the goal is sensitivity or specificity maximisation.

Adenosine deaminase test comparisons

From the unpaired test comparisons, the ADA group appears to be superior in accuracy to all of the other test groups except for the cytokine tests, which appear almost equivalent in accuracy, although the result is significant only for the comparison with anti-TB antibody tests ($p = 0.02$) and antigen tests ($p = 0.02$). The available paired data disagree with this position for the comparison with the NAAT tests (discussed above) and for the comparison with cytokine tests.

The paired adenosine deaminase versus cytokine test comparisons suggest that the ADA tests ($n = 9$) are more accurate than the cytokine tests ($n = 8$); however, on closer inspection of individual test comparisons, it appears that although this is true for the groups overall, the ADA test is actually less accurate than interferon-γ cytokine test (unpaired ROR for interferon-γ in comparison with ADA 15.68, $p < 0.01$; paired ROR 8.33, $p = 0.06$).

Cytokine test comparisons

As a group, the cytokine tests are more accurate than anti-TB antibody tests, antigen tests and lysozyme tests, equivalent to the ADA tests and less accurate than the NAAT tests. However, when the data for the interferon-γ cytokine tests are examined separately from the rest, it actually appears to be more accurate than the NAAT tests and the ADA test, but there are only five studies overall, only one of which used a combined reference standard and none using blinded test interpretation, therefore further research is required to evaluate fully its place for diagnosis of pleural TB infection.

Summary

Overall ADA tests, interferon-γ and NAAT tests appear to perform with approximately equivalent performance when one considers the likely limitations of study design for different tests, although owing to insufficient numbers of studies the analyses do not specifically take account of study design. In particular, none of the ADA evaluations or the cytokine evaluations were considered to have recruited a representative sample of patients, whereas eight of the 20 NAAT evaluations did (*Table 30*). The non-NAAT test evaluations appeared more likely to have simply recruited a group of patients with pleural effusion rather than patients specifically suspected of having TB. NAAT test studies were also more likely to be prospective in design and to report blinded test interpretation. At the same time, the NAAT test evaluations were on average smaller than the other test groups (*Table 30*), over one-third used culture alone as the reference standard and 20% were laboratory based. It is impossible to

know the extent to which the combination of these factors will have affected test accuracy.

Discussion: test accuracy in pleural TB infection

Our review of tests for the detection of pleural TB infection reveals a different picture to that for pulmonary TB infection. In the first instance, the NAAT tests do not perform as well for pleural TB detection as they did for pulmonary infection (DOR 33.18 compared with 116.58). The main problem appears to be with the sensitivity of the tests, presumably due to difficulties in obtaining plural fluid samples with sufficient mycobacteria to allow amplification. Tests that do not rely on detection of mycobacteria, especially ADA tests and tests measuring interferon-γ levels, appear to have performed very well in comparison (DOR 78.54 and 817.48), although the body of evidence for the latter test is very small.

Similarly to the pulmonary TB studies, the quality of reporting is poor, with a large proportion of studies not reporting blinding, study design or sufficient data on which to judge the representativeness of the patients included. However, as discussed in the previous section, the quality of the available studies does vary between groups. The NAAT test studies on average were smaller and were more likely to report the presence of key study design features such as blinding.

Overall, the sensitivity and specificity of the NAAT tests were found to be 72.6 and 93.7%, respectively. A previous meta-analysis of NAAT tests in pleural TB by Pai and colleagues,[59] using an alternative methodology to that which we adopted and including case–control studies and per specimen data, estimated the sensitivity and specificity of the commercial tests at 62 and 98%, respectively (based on 14 studies). We found no significant differences in accuracy between the commercial and in-house tests, and in fact found in-house tests using the IS6110 target sequence to have comparable test accuracy to that of the AMTD test. Pai and colleagues recommended that commercial NAAT tests do have a role in confirming pleural TB, a recommendation which our results confirm (specificity 97.1%).

The evidence for ADA tests, however, suggests that when performed with pleural fluid samples they are at least as reliable at ruling in disease as the NAAT tests and may be of use in ruling out TB infection (specificity 88.4% and sensitivity 92.9%). Paired data from two studies did suggest that in a direct comparison the NAAT tests were more accurate than ADA, but this was predominantly influenced by a single, small study using an inadequate reference standard, which produced very low sensitivity and specificity estimates for ADA.[212]

Two previous meta-analyses assessing ADA for detection of pleural infection also found ADA to perform well (Q-points of 93% from 31 studies[86] and 92.2% from 40 studies[189]). Both reviews included case–control studies and per specimen data. The corresponding Q-point (point of maximum joint sensitivity and specificity) from our analysis discussed above was 89.4%. Both authors conclude, as we do, that ADA estimation has a potential role in the diagnosis of pleural TB infection.

There is also some suggestion for a potential role for interferon-γ or lysozyme measurement, but the available data are much more limited. In particular, interferon-γ, based on five studies, had a summary sensitivity of 94.5% and a specificity of 97.3%. The lysozyme studies, especially those estimating the pleural/serum ratio, gave sensitivity and specificity of 91.5 and 90.6%, respectively. The place of these tests in the diagnosis of pleural TB requires further investigation.

Chapter 8

Results: detection of tuberculous meningitis

Nucleic acid amplification tests

We identified a total of 26 NAAT test evaluations in TB meningitis (*Table 47*). ROC plots of each sensitivity and specificity pairs for each group of tests are provided in *Figure 14(a–e)*.

Description and quality of included studies

The 26 comparisons were carried out in 21 separate studies, all using all CSF specimens. A summary of key characteristics across all data sets is given in *Table 48*, with details per study provided in Appendix 16 grouped according to evaluation of commercial or in-house tests.

Tests

Half of the NAAT evaluations were of commercial tests and half of in-house tests (*Table 48*). Of the in-house tests, 53% (n = 7) used the IS6110 target sequence and the others each used different target sequences including IS986, MPB64, and TRC4; three studies did not report the target sequence used. Of the commercial evaluations, six data sets evaluated the AMTD test, five used Amplicor and LCx and Amplicis Myco B were each evaluated in one study.

TABLE 47 *TB meningitis studies – summary of data sets identified*

	All (CSF)
Total	42
NAAT tests	**26**
Commercial tests	13
AMTD	6
Amplicor	5
LCx	1
Amplicis Myco B	1
In-house tests	13
IS6110	7
Other targets	6
Adenosine deaminase tests	**8**
ADA	7
ADA2	1
Anti-TB antibody tests	**4**
In-house tests	4
BCG	2
H37Rv	2
Antigen tests	**2**
Other miscellaneous tests	**2**
Bromide partition test	2

Reference standards

A combined reference standard of culture plus clinical diagnosis was used in 54% (14/26) of studies (*Table 48*). Culture alone was the reference standard used in a further 42% (11/26). The remaining study did not include culture as part of the reference standard but relied on clinical diagnosis, histology and an anti-TB therapy trial.

Sample details

About 58% (15/26) of studies were hospital based and the remainder laboratory based. All samples tested in these studies were CSF samples.

Overall, the mean number of patients recruited was 89 (SD 89; range 6–402) and there was little difference between the numbers of patients evaluated in in-house and commercial studies, although the range was wider for the commercial studies (*Table 48*). The mean prevalence of TB was 29% (SD 19%).

We judged the patient sample to have been representative (i.e. the study at least stated that that patients 'suspected' of having TB were recruited) in 77% (20/26) of studies. The sample was judged to be clearly unrepresentative in two studies and was unclear in the remaining four.

Test interpretation

Index test interpretation was reported as blinded in 27% (7/26) studies and was not clearly reported in the rest. Nine studies were also clearly prospective in design and two were retrospective. Assuming (based on time taken to perform PCR compared with culture) that the index test was interpreted first in the prospective studies and was therefore to all intents and purposes 'blinded', increased the number of studies in which the index test was potentially interpreted blinded to 14 (54%). The reference standard was clearly reported as interpreted blinded in two studies (all of which also reported blinded index test interpretation), was not blinded in two studies and not clearly reported in the remainder (*Table 48*).

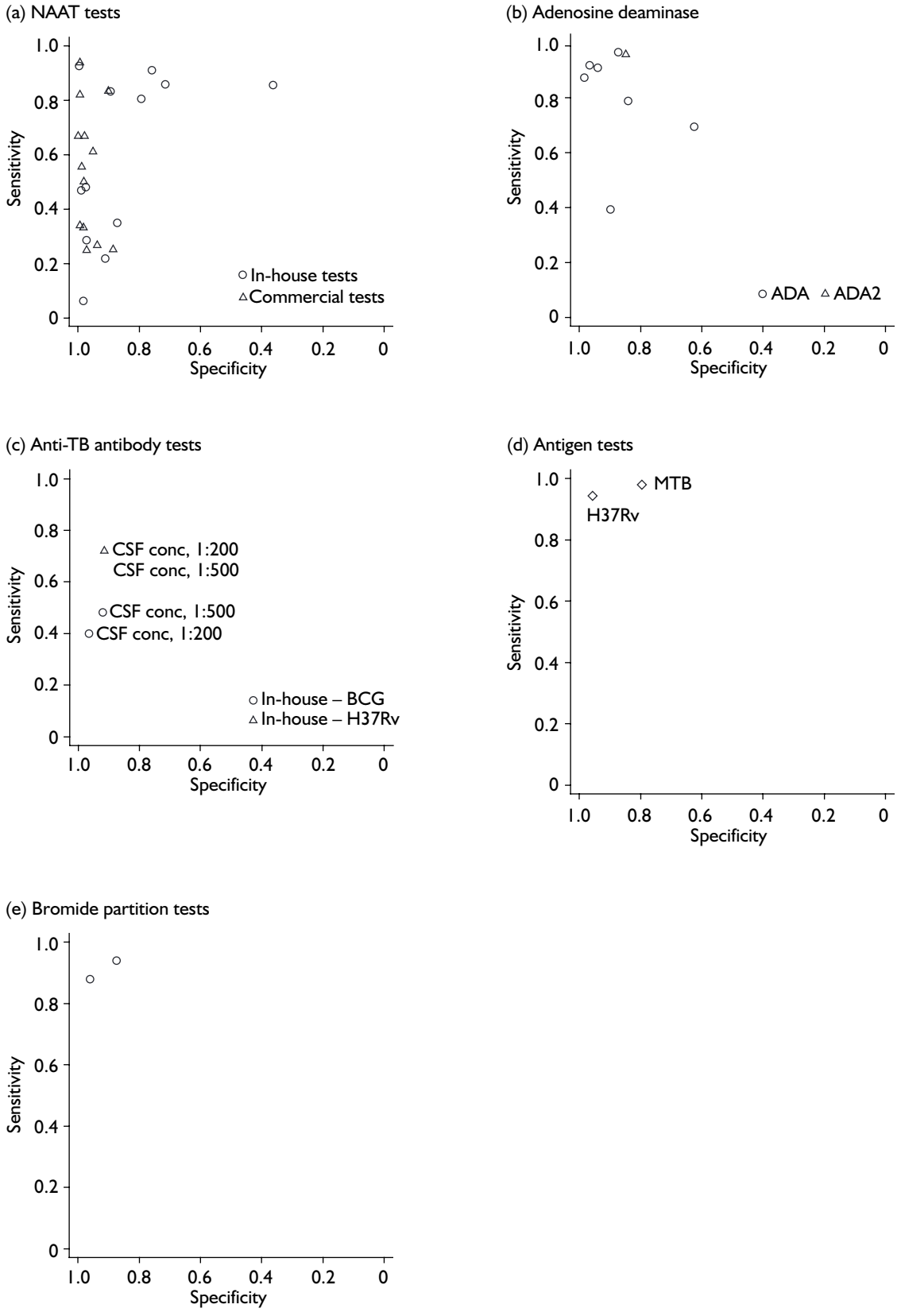


FIGURE 14 *TB Meningitis – ROC plots for each group of tests*

TABLE 48 *Diagnostic test evaluations in TB meningitis – summary of key study characteristics for main tests*

	NAAT	Adenosine deaminase
Total no. of studies	26	8
Mean sample size	89 (SD 89; range 6–402)	86 (SD 39; range 33–136)
Commercial	13	
Mean sample size	97 (SD 120; range 6–402)	
In-house	13	8
Mean sample size	82 (SD 42; range 15–146)	
Reference standard		
Culture + clinical diagnosis ± other	14 (54%)	5 (63%)
Culture + anti-TB therapy	0	1 (13%)
Culture + other	0	0
Culture alone	11 (42%)	2 (25%)
No culture	1 (4%)	0
Disease prevalence (mean, SD, range)	29% (SD 19; range 1–70%)	29% (SD 21; range 4–60%)
Setting		
Hospital-based	15 (58%)	8 (100%)
Clinic	0	0
Laboratory-based	8 (31%)	0
Unknown	0	0
Sample type		
CSF	26 (100%)	8 (100%)
Patients representative?		
Yes	20 (77%)	0
No	2 (8%)	5 (63%)
Unclear	4 15%)	3 (38%)
Study design prospective?		
Yes	9 (35%)	3 (38%)
No	2 (8%)	1 (13%)
Unclear	15 (58%)	4 (50%)
Index test interpreted blinded?		
Yes	7 (27%)	3 (38%)
No	0	0
Unclear	19 (73%)	5 (63%)
Reference test interpreted blinded?		
Yes	2 (8%)	1 (13%)
No	2 (8%)	0
Unclear	22 (85%)	7 (88%)

TABLE 49 *NAAT evaluations in TB meningitis – tests for heterogeneity of sensitivities and specificities*

		Sensitivity		Specificity	
Comparison	***n***	***Q statistic***	***p*-Value**	***Q statistic***	***p*-Value**
All NAAT comparisons	26	124	<0.01	143	<0.01
Commercial tests	**13**	**26**	**0.01**	**17**	**0.15**
AMTD	6	17	<0.01	4	0.51
Amplicor	5	3	0.54	7	0.12
In-house tests	**13**	**97**	**<0.01**	**67**	**<0.01**
IS6110	7	46	<0.01	17	0.01
Other targets	6	51	<0.01	35	<0.01

Results

Sensitivity and specificity

The plot of sensitivity against specificity for each study is presented in *Figure 14(a)*. Overall, sensitivity and specificity are highly variable; the tests for heterogeneity for both indices were statistically significant at $p < 0.01$ (*Table 49*). When studies are split according to whether they are commercial or in-house tests, the ROC plots and heterogeneity tests (*Figure 15* and *Table 49*) indicate that sensitivity varies for both groups ($p < 0.01$), but specificity estimates are far less

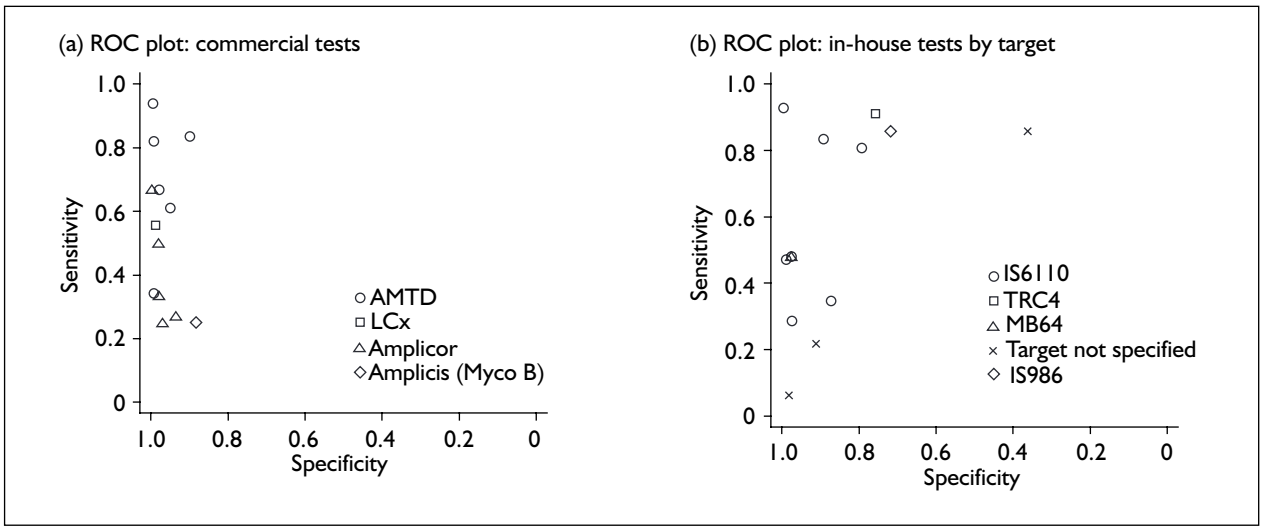


FIGURE 15 *TB Meningitis – ROC plots for NAAT evaluations by test or target*

TABLE 50 *NAAT evaluations in TB meningitis – regression analyses to identify source(s) of heterogeneity*

Comparison	*n* (26)	Model parameters (95% CI) and *p*-value		
		Coefficient	*p*-Value	RDOR[a]
Test type				
AMTD vs in-house	6/13	1.88	0.06	6.53 (0.94 to 45.29)
Amplicor vs in-house	5/13	0.38	0.73	1.46 (0.16 to 13.41)
LcX vs in-house	1/13	1.39	0.49	4.01 (0.07 to 242.79)
Amplicis Myco B vs in-house	1/13	–2.17	0.28	0.11 (0.00 to 6.54)
Institute				
Laboratory-based vs hospital-based	8/18	2.32	<0.01	10.19 (2.39 to 43.33)
Reference standard				
Culture plus clinical vs culture alone	14/11	–0.05	0.95	0.95 (0.18 to 5.05)
No culture vs culture alone	1/11	–2.51	0.31	0.08 (0.00 to 11.54)
Design				
Prospective vs retrospective/unknown	9/17	–0.37	0.69	0.69 (0.10 to 4.61)
Quality factors				
Patients representative vs unrepresentative/ unknown	20/6	0.09	0.92	1.09 (0.16 to 7.54)
Index test blinded vs not blinded/unknown	7/19	–1.42	0.11	0.24 (0.04 to 1.38)
Reference test blinded vs not blinded/unknown	2/24	–1.80	0.23	0.17 (0.01 to 3.43)
TB prevalence				
Prevalence <10% vs prevalence 30+%	5/11	1.74	0.08	5.72 (0.80 to 41.15)
Prevalence 10–20% vs prevalence 30+%	4/11	1.96	0.07	7.11 (0.83 to 60.57)
Prevalence 20–30% vs prevalence 30+%	6/11	2.56	0.01	12.97 (1.86 to 90.52)
% patients smear positive				
<5% smear positive vs >30%	8/5	–0.52	0.70	0.59 (0.03 to 10.43)
10–20% smear positive vs >30%	3/5	–1.03	0.56	0.36 (0.01 to 14.58)

[a] Relative diagnostic odds ratio, i.e. in one group compared with the other.

variable for the commercial tests and the heterogeneity test becomes non-significant ($p = 0.15$). The statistical tests for heterogeneity for individual commercial tests suggest that sensitivity estimates for the Amplicor test vary less than for the AMTD test; however, the number of studies is small ($n = 5$) and it can be seen from *Figure 15(a)* that sensitivity still ranges from just over 20 to almost 70%.

Both sensitivities and specificities of the in-house tests vary considerably and heterogeneity remains when subdivided by the target sequence used (IS6110 compared with other).

Heterogeneity investigation

The heterogeneity investigations (*Table 50*) indicate that the only design-related variable to have a significant impact on test accuracy was whether studies were hospital based or not: accuracy was more than 10 times greater in laboratory-based than hospital-based studies (RDOR 10.19; 95% CI: 2.39 to 43.33). Reference standard used and prospective/retrospective design had no impact on accuracy. Test performance was lower in studies using index or reference test blinding, but neither of these reached statistical significance. Accuracy again was higher in lower prevalence studies and the difference reached statistical significance for one comparison (*Table 50*).

SROC analysis

The overall pooled analysis suggests that the NAAT tests have high specificity in CSF specimens but generally poor sensitivity, regardless of test (*Table 51*). Individually, the AMTD test appears to perform the best (sensitivity 74% and specificity 98%). The other commercial tests demonstrated consistently low sensitivity. Neither the sensitivity nor specificity of the inhouse tests in general was sufficiently high to justify their recommendation for use in diagnosis of TB meningitis. None of the 26 studies met all five study design-related criteria.

Summary

Overall, the studies in this section were not well reported, especially in terms of study design and blinding of test interpretation. This, along with the small number of studies included, may have made it difficult to detect any impact on test accuracy of the various study design features. The AMTD test was considerably superior to the others, especially in terms of its specificity. NAAT tests provide a reliable means of ruling in the presence of TB meningitis.

Serodiagnostic and biochemical tests for detection of tuberculous meningitis

Description and quality of included studies

We included 16 data sets relating to other tests that have been used to detect the presence of TB infection (*Table 47*). A summary of study characteristics for the ADA-related tests is provided in *Table 48*. Summary details per study are provided in Appendix 17.

Adenosine deaminase tests

The eight ADA tests were evaluated in seven studies (Appendix 17); all evaluated ADA with one also assessing ADA2[214] for the detection of TB infection. The mean sample size was 86 (SD 39%) and TB prevalence was 29% (SD 21%) (*Table 48*). All studies were hospital based; five used a combined culture plus clinical diagnosis reference standard, two used culture alone and one used culture plus histology. The study design was clearly prospective in four studies, retrospective in one and could not be determined for the others. Index test interpretation was blinded in three studies, one of which also reported blinded reference test interpretation. The patient samples were judged to be unrepresentative in five of the eight evaluations and were unclear for the others.

Anti-TB antibody tests

The four anti-TB antibody tests evaluations were all carried out in a single study[215] evaluated in two studies, each evaluating two tests in two different samples (Appendix 17). Two antigens were used (BCG and H37Rv) and each was evaluated in two different concentrations of CSF (1:200 and 1:500). The study sample size was relatively large (n = 260) and TB prevalence was 10%. A combined culture plus clinical diagnosis reference standard was used and index test interpretation was blinded. None of the other design-related characteristics could be assessed.

Antigen tests

Two antigen tests were evaluated in two separate studies (Appendix 17). Both were relatively small in size (n = 41 and 53) and we were not able to determine whether key study design features were present.

Other miscellaneous tests

The bromide partition test was assessed in two studies. Both were retrospective in design, with less than 100 patients each (Appendix 17). We could not determine whether blinded test interpretation was implemented. Patients were judged to be representative in one of the studies.

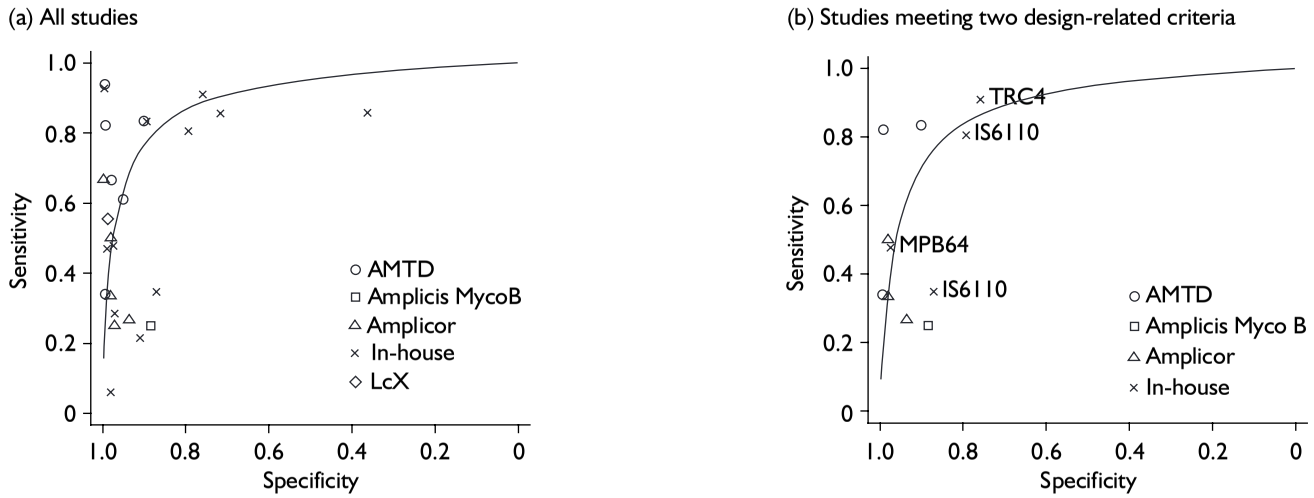
Results

The plot of sensitivity against specificity for each group of tests is presented in *Figure 14(b–e)* and results of the tests for heterogeneity for both indices in *Table 52*.

Adenosine deaminase tests

The seven data sets were very heterogeneous for both sensitivity and specificity ($p < 0.01$). The overall summary DOR was 65.63 (95% CI: 11.39

TABLE 51 *NAAT evaluations in TB meningitis – SROC regression models and associated SROC curves*



Comparison	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
All tests	26	26.72 (7.15 to 99.80)	0.58	58.6	96.2	11	20.87 (3.30 to 131.90)	0.83	56.1	95.0
All commercial	**13**	**67.29 (1.72 to 2633.93)**	**0.94**	**56.4**	**97.9**	**7**	**29.32 (0.19 to 4511.07)**	**0.98**	**48.8**	**97.0**
AMTD	6	146.98 (1.34 to 16079.24)	0.99	74.1	98.0	3	115.63 (3.92 to 3409.97)	NA	69.4	98.1
Amplicor	5	35.67 (3.24 to 392.45)	NA	39.6	98.2	3	18.75 (1.07 to 327.96)	NA	36.2	97.1
LcX	1	96.25 (8.99 to 1030.34)	NA	55.6	98.7	0				
Amplicis Myco B	1	2.56 (0.07 to 95.88)	NA	25.0	88.5	1	2.56 (0.07 to 95.88)	NA	25.0	88.5
In-house	**13**	**19.44 (4.19 to 90.24)**	**0.86**	**60.7**	**93.2**	**4**	**15.87 (3.02 to 83.50)**	**NA**	**68.1**	**88.2**
Subgroups meeting each design-related criterion										
Hospital-based	18	15.68 (4.57 to 53.76)	0.81	48.8	94.8					
Combined reference test	14	28.66 (3.49 to 235.35)	0.73	58.9	96.4					
Index test blinded/prospective	14	31.75 (2.62 to 385.05)	0.76	47.2	96.3					
Reference test blinded	2	6.21 (0.00 to 497,494.61)	NA	58.6	81.5					
Patients representative	20	27.00 (5.19 to 140.37)	0.63	57.8	96.4					
Subgroups by smear status										
Smear negative	2	185.36 (0.00 to 191,387,360.94)	NA	10.7	90.5					

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

TABLE 52 *Serodiagnostic and biochemical test evaluations in TB meningitis – tests for heterogeneity of sensitivities and specificities*

		Sensitivity		Specificity	
Comparison	***n***	***Q statistic***	***p*-Value**	***Q statistic***	***p*-Value**
Adenosine deaminase tests	8	25	<0.01	27	<0.01
Anti-TB antibody tests	4	8	0.05	6	0.12
Antigen tests	2	0	0.63	3	0.07
Bromide partition tests	2	1	0.32	1	0.29

to 378.22) with associated sensitivity of 86.5% and specificity of 90.5% (*Table 53*). Four of the eight studies (all using ADA) were both hospital based and used combined reference standards; test accuracy in this group increased to give a summary sensitivity of 87% and specificity of 88%. None of the studies met all five design-related criteria.

Anti-TB antibody tests

We did not carry out any pooling of the anti-TB antibody data sets as they all came from the same study. *Figure 14(c)* shows that regardless of CSF concentration used, the presence of anti-TB antibodies to the $HR37R_v$ antigen were more sensitive but less specific than the BCG antigen for the diagnosis of TB infection.

Antigen tests

The two antigen tests evaluated were found to have high sensitivity and specificity in two separate studies (*Table 54*). The combined result gives a sensitivity of 96.5% and a specificity of 90.2%. Both studies met our first two design-related criteria, but none met all five.

Other miscellaneous tests

The two bromide partition test evaluations identified also had fairly high sensitivity and specificities (*Table 55*). Both studies were hospital based and used a combined reference standard; summary sensitivity was 91.5% and specificity 92.8%. Neither met all five design-related criteria.

Summary

Three groups of tests demonstrated reasonably high sensitivity and specificity for the detection of TB meningitis: ADA tests, antigen tests and bromide partition tests. However, many of the studies in these groups had small sample sizes, none reported having used blinded test interpretation and in only one[216] was the patient sample judged to be representative. These factors combined mean that the evidence for the accuracy of these tests in detecting TB infection is not yet proven.

Empirical comparisons between test types

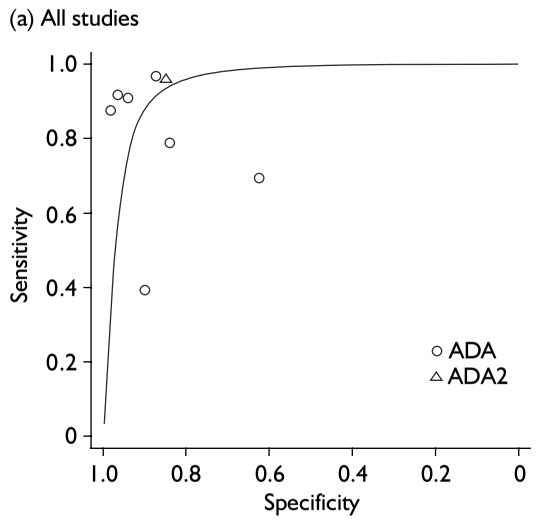
For the TB meningitis test comparisons (*Table 56*), the unpaired comparisons suggest that ADA tests were more accurate than the NAAT tests and both were apparently less accurate than the antigen tests; none of the comparisons were statistically significant. However, this does not take into account the effect of any potential confounding on accuracy. For example, similarly to the pleural TB studies, none of the adenosine deaminase data sets were from patient samples that were judged to be representative, whereas 20 of the 26 NAAT test evaluations were representative. Other factors including blinding of test interpretation and reference standard used may also considerably affect the overall comparison. Nevertheless further research is needed to establish the role of adenosine deaminase in the diagnosis of TBM.

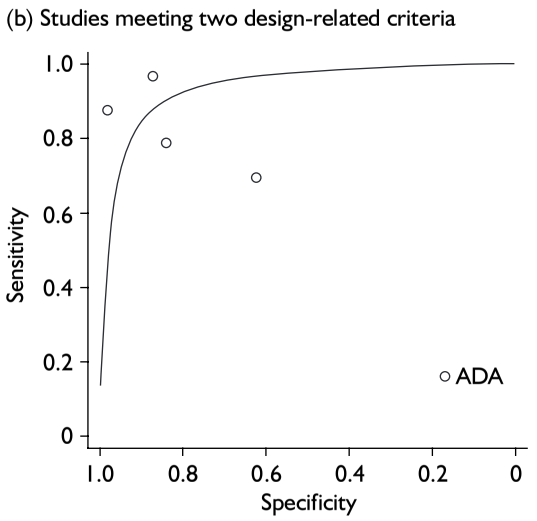
The two antigen tests also appear superior in accuracy to both the NAAT tests and the ADA tests, but both studies included few patients (less than 55 each) and study design details were not clear.

Discussion: test accuracy in TB meningitis

The majority of the data for tests for TB meningitis are for the use of NAAT tests where a similar picture to that for pleural TB emerges (DOR 26.72 compared with 33.18). Again, the problem appears to be with the low sensitivity of the NAATs, also presumably due to the mycobacterial load typically present in CSF samples in TB meningitis. Again, the ADA tests, which do not rely on direct detection of mycobacteria, have much higher accuracy (DOR 65.63), although the body of evidence is very small. The unpaired comparisons between groups suggest that the ADA tests were considerably more accurate than the NAAT tests.

TABLE 53 *Adenosine deaminase tests in TB meningitis – SROC regression models and associated SROC curves*



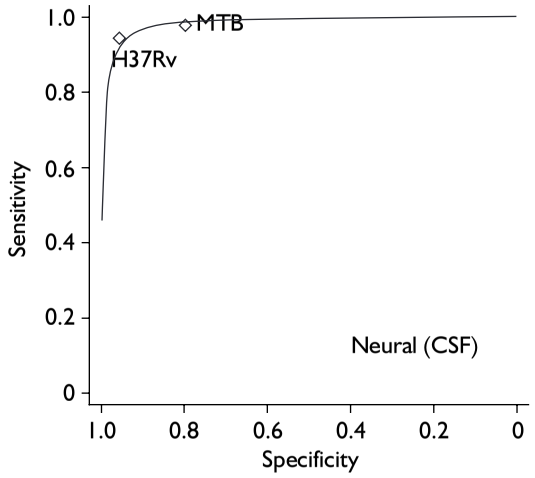


Comparison	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
Adenosine deaminase tests[e]	8	65.63 (11.39 to 378.22)	0.71	86.5	90.5	4	48.61 (1.65 to 1432.21)	NA	86.5	88.3
ADA	7	59.17 (5.83 to 600.24)	0.86	84.2	91.1	4	As above			
ADA2	1	128.14 (6.06 to 2709.93)	NA	95.8	84.8					

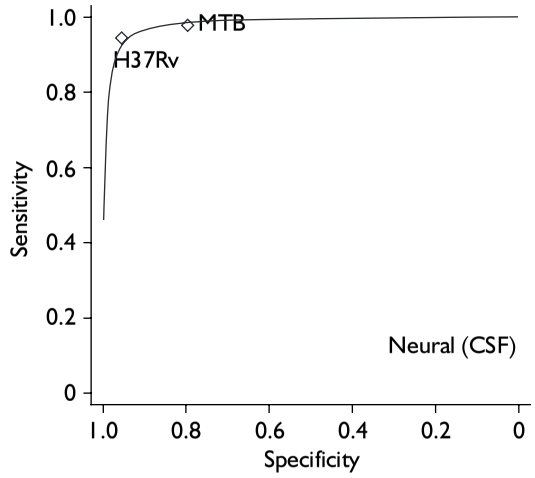
[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.
[e] Includes three pleural TB data sets, one peritoneal and two in pulmonary TB.

TABLE 54 *Antigen tests in TB meningitis – SROC regression models and associated SROC curves*

(a) All studies



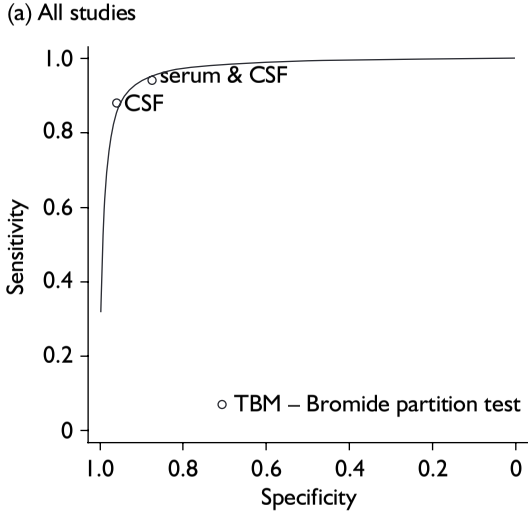
(b) Studies meeting two design-related criteria

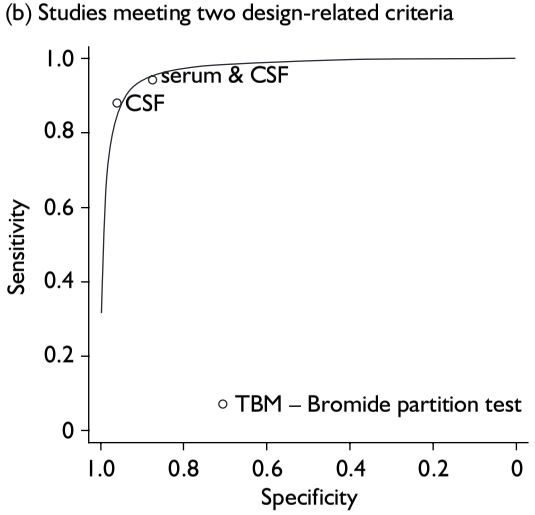


Comparison	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
TB meningitis	2	255.00 (2.38 to 27,272.93)	NA	96.5	90.2	2	255.00 (2.38 to 27,272.93)	NA	96.5	90.2

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

TABLE 55 *Miscellaneous other tests in TB meningitis – SROC regression models and associated SROC curves*





Comparison	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
Bromide partition test[e]	2	139.60 (8.50 to 2293.13)	NA	91.5	92.8	2	139.60 (8.50 to 2293.13)	NA	91.5	92.8

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.
[e] Includes three pleural TB data sets, one peritoneal and two in pulmonary TB.

TABLE 56 *TB meningitis – comparisons between test types*[a]

		NAATs (*n* = 26)			**Adenosine deaminase (*n* = 8)**	
Adenosine deaminase	26/8	ROR = 2.02, $p = 0.43$	←			
	0/0	–				
Antigen test	26/2	ROR=10.85, $p = 0.14$	←	8/2	ROR = 3.59, $p = 0.44$	←
	0/0	–		0/0	–	

ROR, relative odds ratio – ratio of DORs.
[a] ↑, ←, arrows point towards test with higher DOR. Upper cells indicate unpaired comparisons; lower cells indicate results from paired comparisons.

Overall, the ADA tests actually appear superior in quality to the NAAT tests. Mean sample sizes were similar but a higher proportion of ADA studies used a combined reference test, were prospective in design and reported blinded test interpretation, although overall these proportions were still low. The patients included in the NAAT test studies were, however, more likely to be considered representative.

Comparing the summary sensitivity and specificity estimates between the two groups of tests shows that although the ADA tests have a better combined sensitivity and specificity (86.5 and 90.5%, respectively), the NAAT tests are greatly superior in terms of specificity (sensitivity 58.6% and specificity 96.2%). This suggests that NAAT tests should in fact be the first-line test for ruling in TB meningitis, but that it needs to be combined with the result of other tests in order to rule out disease. This conclusion is similar to that of a previous meta-analysis despite differences in study inclusion and review methodology.[56]

Surprisingly, two other tests (antigen tests and the bromide partition test), both evaluated in two studies, were each found to have high sensitivity and specificity. However, both were small in size, and it is not clear whether these tests are worthy of further investigation.

The anti-TB antibody tests performed poorly in the single study included.

Chapter 9

Results: detection of lymphatic tuberculosis infection

We identified 18 test comparisons in lymphatic TB (*Table 57*). ROC plots of each sensitivity and specificity pair for each group of tests are provided in *Figure 16(a–c)*.

Nucleic acid amplification tests

Sixteen NAAT test evaluations in pleural TB carried out in 14 separate studies were included. A summary of key characteristics across all data sets is given in *Table 58*, with details per study provided in Appendix 18 grouped according to evaluation of commercial or in-house tests.

Description and quality of included studies

Tests

More than half of the NAAT evaluations (9/16) were of commercial tests; four of AMTD, four Amplicor and one of the LCx test. Of the in-house tests, three (43%) used the IS6110 target sequence, two used 65 kDa and the others used MPB64 and TRC4 (Appendix 18).

Reference standards

A combined reference standard of culture plus clinical diagnosis was used in only one study (*Table 58*). Culture alone was relied upon in seven studies (44%) and culture plus anti-T therapy treatment, histology or X-ray in six studies. The remaining two did not include culture as part of the reference standard but relied on cytology alone[34] or histology plus an anti-TB therapy trial.[217]

TABLE 57 *Lymphatic TB studies – summary of data sets identified*

	All
TOTAL	18
NAAT tests	**16**
Commercial tests	9
AMTD	4
Amplicor	4
LCx	1
In-house tests	7
IS6110	3
Other targets	4
Anti-TB antibody tests	**1**
In-house test	1
H37Ra	1
Antigen tests	**1**
MTB	1

Sample details

About 75% (12/16) of studies were hospital based, three were laboratory-based and in one the setting was not clear. All samples tested in these studies were lymph node biopsy specimens.

Overall, the mean number of patients recruited was 50 (SD 31; range 4–113) and there was very little difference between the numbers of patients evaluated in in-house compared with commercial studies (*Table 58*). The mean prevalence of TB was 53% (SD 15%).

We judged the patient sample to have been representative (i.e. the study at least stated that that patients 'suspected' of having TB were recruited) in 75% (12/16) of studies.

Test interpretation

Index test interpretation was reported as blinded in one study[218] and was not clearly reported in the rest. None of the studies appeared to be prospective in design and three were retrospective. The reference standard was clearly reported as interpreted blinded in one study, was not blinded in one and not clearly reported in the remainder (*Table 58*).

Results

Sensitivity and specificity

The plot of sensitivity against specificity for each study is presented in *Figure 16(a)* and statistical tests for heterogeneity in *Table 59*. Overall, specificity in particular is highly variable ($p < 0.01$), but sensitivity estimates are less so ($p = 0.05$). When studies are split according to whether they are commercial or in-house tests, the ROC plots (*Figure 17*) demonstrate that it is the specificity of the in-house tests that appears to be causing much of the variation. In particular, two of the IS6110 studies and the study using TRC4 as the target sequence have especially low specificity.

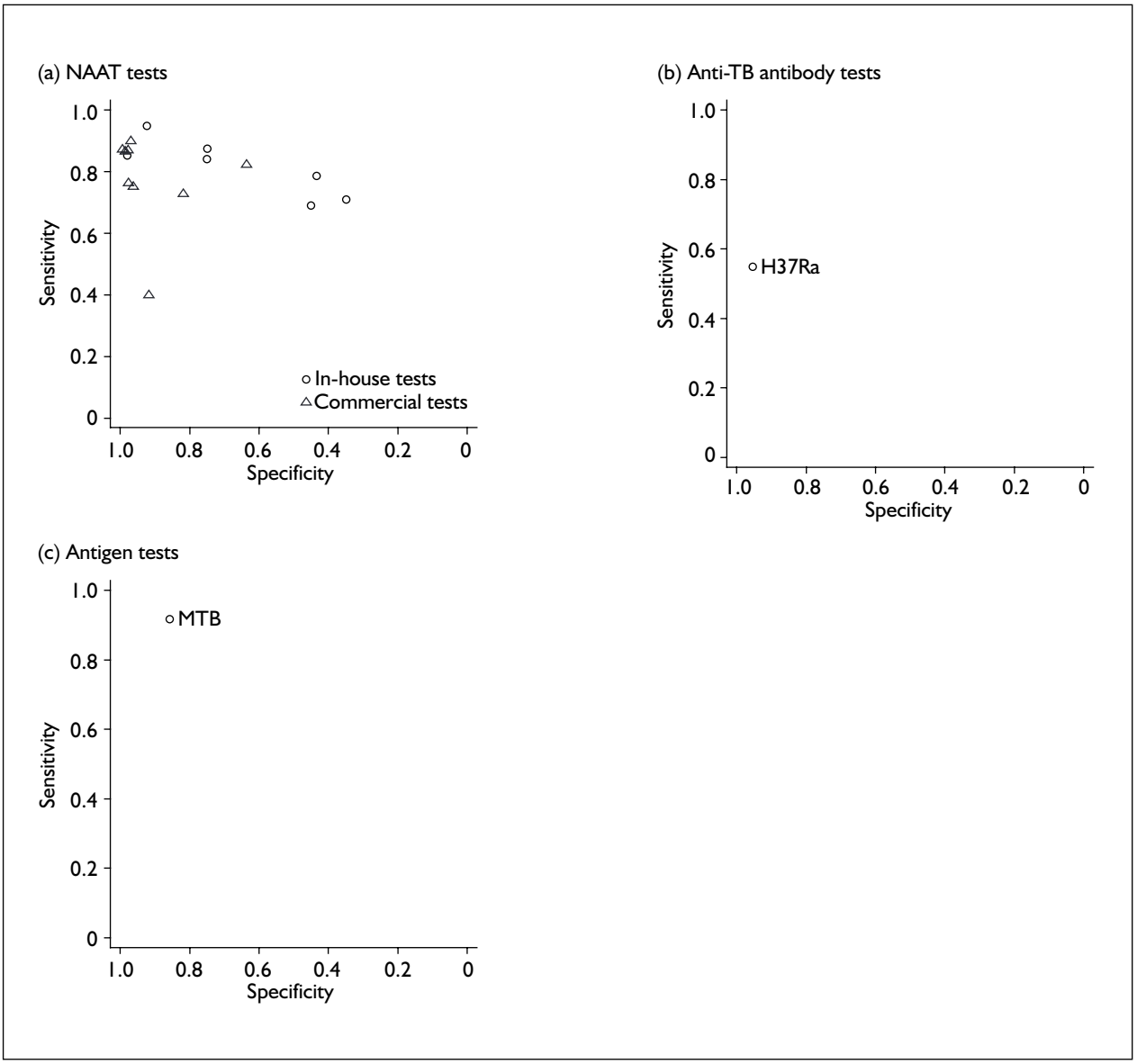


FIGURE 16 *Lymphatic TB – ROC plots for all test types*

Heterogeneity investigation

None of the variables investigated had any significant impact on test accuracy (*Table 60*).

SROC analysis

The overall pooled analysis suggests that the NAAT tests have relatively high specificity in lymph node specimens but sensitivity is less good (*Table 61*). Specificity was generally higher for the commercial tests (DOR 21.82; 95% CI: 2.28 to 208.67; sensitivity 79.6%, specificity 95.7%) compared with the in-house tests (DOR 30.77; 95% CI: 6.23 to 152.01; sensitivity 83.3, specificity 74.5). This was primarily due to the AMTD and LCx tests, which performed much better than the Amplicor test.

Only one study with only four patients was hospital based and used a combined reference standard.[219] None of the 16 studies met all five study design-related criteria.

Summary

Overall, the studies in this section were very poorly reported, especially in terms of study design and blinding of test interpretation. This, along with the small number of studies included, may have made it difficult to detect any impact on test accuracy of the various study

TABLE 58 *NAAT evaluations in lymphatic TB – summary of key study characteristics*

	All studies
Total no. of studies	16
Mean sample size	50 (SD 31; range 4–113)
Commercial	9
Mean sample size	44 (SD 27; range 28–113)
In-house	7
Mean sample size	57 (SD 36; range 4–101)
Reference standard	
Culture + clinical diagnosis ± other	1 (6%)
Culture + anti-TB therapy	2 (13%)
Culture plus other (histology, X-ray, etc.)	4 (25%)
Culture alone	7 (44%)
No culture	2 (13%)
Disease prevalence (mean, SD, range)	53% (SD 15; range 22–76%)
Setting	
Hospital-based	12 (75%)
Laboratory-based	3 (19%)
Unknown	1 (6%)
Patients representative?	
Yes	12 (75%)
No	0
Unclear	4 (25%)
Study design prospective?	
Yes	0
No	3 (19%)
Unclear	12 (75%)
Index test interpreted blinded?	
Yes	1 (6%)
No	0
Unclear	15 (94%)
Reference test interpreted blinded?	
Yes	1 (6%)
No	1 (6%)
Unclear	14 (88%)

design features. There is currently insufficient evidence on which to recommend use of this test.

Other serodiagnostic and biochemical tests

Description and quality of included studies

We included two data sets relating to serodiagnostic and other tests to detect lymphatic TB infection (*Table 57*). Summary details per study are provided in Appendix 19.

Anti-TB antibody tests

The anti-TB antibody test study used the H37Ra antigen to detect anti-TB antibodies.[220] A total of 126 patients were included and TB prevalence was 23%. The reference standard was culture plus histology. We judged the patients sample to be representative but could not determine the study design or whether the test interpretation had been blinded.

Antigen tests

The antigen test aimed to detect the *M. TB* antigen.[221] A total of 124 patients were included and TB prevalence was 77%. The reference standard was culture alone. Again, we judged the patients sample to be representative but could not determine the study design or whether the test interpretation had been blinded.

Results

The plot of sensitivity against specificity for each group of tests is presented in *Figure 16(b)* and *(c)* and results in Appendix 19.

TABLE 59 *NAAT evaluations in lymphatic TB – tests for heterogeneity of sensitivities and specificities*

		Sensitivity		Specificity	
Comparison	***n***	***Q statistic***	***p*-Value**	***Q statistic***	***p*-Value**
All NAAT comparisons	16	25	0.05	72	<0.01
Commercial tests	9	15	0.07	18	0.02
AMTD	4	1	0.73	9	0.03
Amplicor	4	7	0.08	4	0.24
LCx	1	NA	NA	NA	NA
In-house tests	7	10	0.11	18	0.01
IS6110	3	2	0.42	9	0.01
Other targets	4	4	0.24	8	0.04

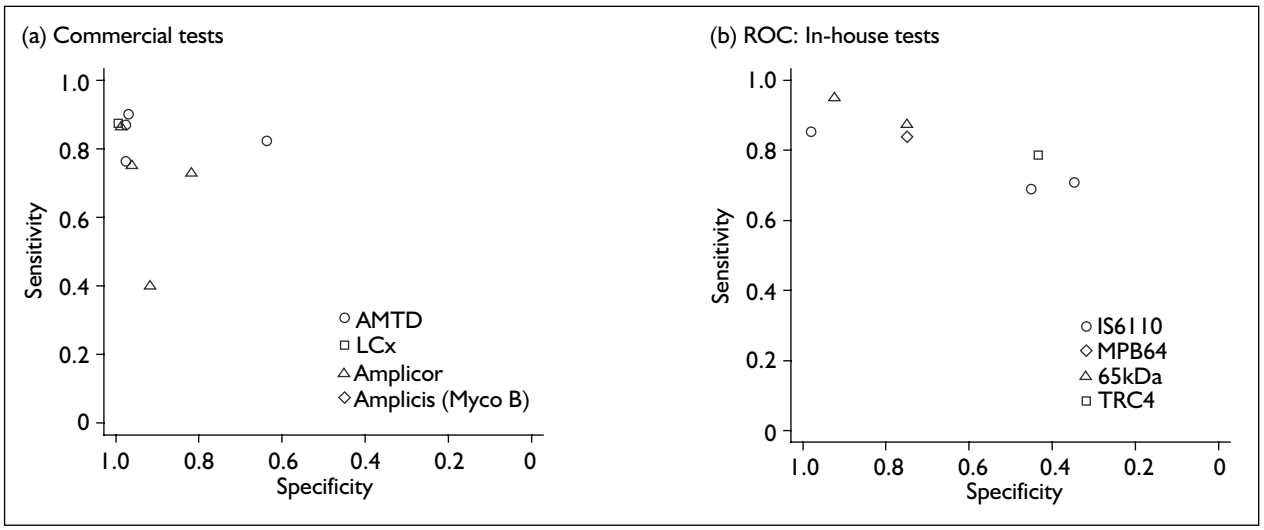


FIGURE 17 *Lymphatic TB – ROC plots for NAAT evaluations by test or target*

TABLE 60 *NAAT evaluations in lymphatic TB – regression analyses to identify source(s) of heterogeneity*

Comparison	*n* (16)	Model parameters (95% CI) and *p*-value		
		Coefficient	*p*-Value	RDOR[a]
Test type				
AMTD vs in-house	4/7	0.03	0.98	1.03 (0.09 to 12.03)
Amplicor vs in-house	4/7	–1.68	0.22	0.19 (0.01 to 3.16)
LcX vs in-house	1/7	0.29	0.89	1.34 (0.01 to 124.56)
Institute				
Laboratory based vs hospital-based	3/13	1.79	0.10	5.98 (0.66 to 54.61)
Reference standard				
Culture plus clinical vs culture alone	1/7	1.00	0.61	2.72 (0.04 to 175.35)
Culture plus other vs culture alone	6/7	0.39	0.71	1.47 (0.16 to 13.50)
No culture vs culture alone	1/7	–0.87	0.54	0.42 (0.02 to 8.58)
Design				
Prospective vs retrospective/unknown	0/16			
Quality factors				
Patients representative vs unrepresentative/ unknown	12/4	–0.41	0.73	0.66 (0.06 to 7.97)
Index test blinded vs not blinded/unknown	1/15	3.09	0.05	21.97 (0.99 to 486.89)
Reference test blinded vs not blinded/unknown	1/15	3.09	0.05	21.97 (0.99 to 486.89)
TB prevalence				
Prevalence 20–30% vs prevalence 30%+	1/15	1.03	0.55	2.80 (0.07 to 104.92)
% patients smear positive				
5–10% smear positive vs 20%+	3/2	–3.37	0.10	0.03 (0.00 to 2.24)
10–20% smear positive vs 20%+	5/2	–1.97	0.23	0.14 (0.00 to 5.05)

[a] Relative diagnostic odds ratio, i.e. in one group compared with the other.

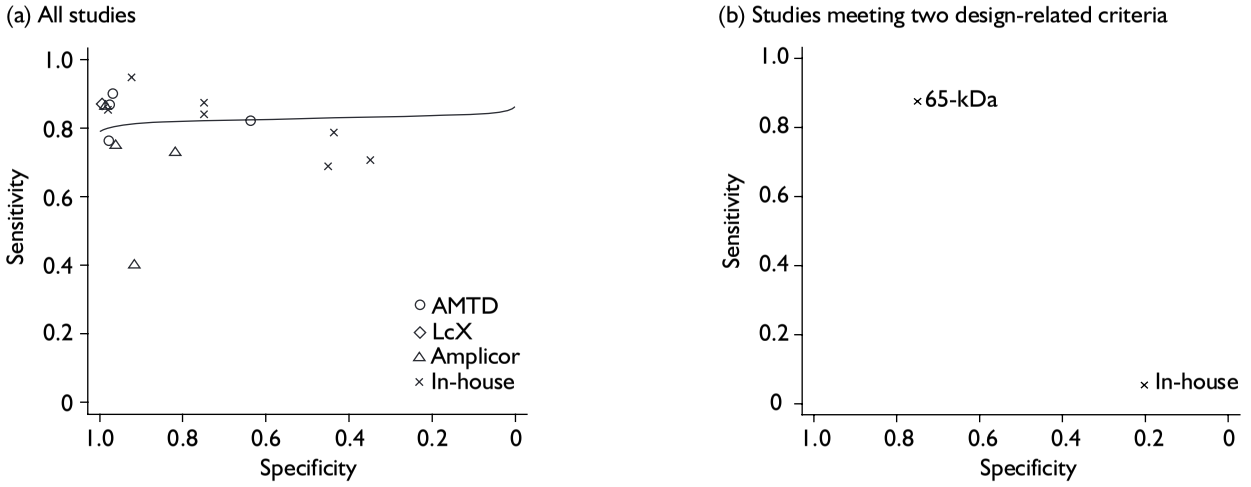
Anti-TB antibody tests

The study demonstrated high specificity (95%) but very low sensitivity (55%) for the detection of TB infection.[220] The sample size was relatively small and it was not possible to judge the quality of the study.

Antigen tests

The study demonstrated high sensitivity (92%) and relatively good specificity (86%) for the detection of TB infection.[221] The sample size was relatively small and prevalence very high and it was not possible to judge the quality of the study.

TABLE 61 *NAAT evaluations in lymphatic TB – SROC regression models and associated SROC curves*



Comparison	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
All tests	16	20.09 (8.16 to 49.46)	<0.01	81.3	90.1	1	21.00 (0.27 to 1646.32)	NA	87.5	75.0
Commercial	9	21.82 (2.28 to 208.67)	0.12	79.6	95.7	0				
AMTD	4	95.00 (6.64 to 1358.78)	NA	84.5	94.6					
Amplicor	4	41.51 (2.09 to 822.95)	NA	70.7	94.5					
LcX	1	1077.67 (56.32 to 20620.10)	NA	87.1	99.4	0				
In-house	7	30.77 (6.23 to152.01)	0.04	83.3	74.5	1	As above			
Subgroups meeting each design-related criterion										
Hospital-based	13	16.51 (6.22 to 43.80)	0.02	79.4	85.0					
Combined reference test	7	28.12 (6.52 to 121.28)	0.06	84.1	84.1					
Index test blinded/ prospective	0									
Reference test blinded	1	222.00 (18.46 to 2670.22)	NA	94.8	92.3					
Patients representative	12	18.90 (5.22 to 68.44)	0.01	80.4	93.9					

[a] Where <5 studies per subgroup (or obtain 'odd'-looking SROC curve), studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

Summary

There is insufficient evidence on which to recommend use of these tests.

Discussion: test accuracy in lymphatic TB

Evidence for the detection of lymphatic TB is significantly less prolific than for the types of TB discussed previously. The largest group of tests are the NAAT tests, for which the summary DOR was 20.09. In this instance, however, it is the specificity of the test that is more variable than the sensitivity, perhaps because the tests were generally performed in lymph biopsy specimens as opposed to body fluid. The summary sensitivity was 81.3% and specificity 90.1%. The in-house tests in this case provided most of the variation in accuracy and also demonstrated the presence of threshold effect (test of symmetry $p = 0.04$). When restricted to commercial tests only, sensitivity was 79.6% and specificty 95.7%.

One anti-TB antibody test was evaluated in lymphatic TB; specificity was found to be high but sensitivity very low at 55%. The antigen test, similarly to the case for TB meningitis, showed surprisingly high accuracy (sensitivity 92%, specificity 86%). It was not possible to judge the quality of these studies and it is difficult to say whether there is any case for further examination of them in future studies.

Chapter 10

Results: detection of peritoneal tuberculosis infection

We identified 14 test comparisons in peritoneal TB (*Table 62*). ROC plots of each sensitivity and specificity pair for each group of tests are provided in *Figure 18(a–d)* and SROC curves for the ADA and cytokine tests in *Figure 19(a, b)*.

Description and quality of included studies

The 14 test evaluations were carried out in seven studies; summary details per study are provided in Appendix 20.

Nucleic acid amplification tests

One study evaluating a NAAT test in ascitic fluid was included[222] (Appendix 20). An in-house test using the IS6110 target sequence was evaluated in seven patients.

Adenosine deaminase tests

Six studies provided eight data sets evaluating the ADA test in ascitic fluid samples ($n = 6$), serum samples ($n = 1$) or both (ascitic/serum ratio, $n = 1$). Three studies (five data sets) used a combined culture plus clinical diagnosis reference standard. Five of the six studies were prospective in design, but only one reported using blinded test interpretation. None of the study samples were judged to have been representative.

Cytokine tests

Two studies evaluated cytokine tests (Appendix 20); one using total protein[223] and one evaluated interferon-γ.[224] Less than 100 patients were included in each study; both were prospective in design but the patient samples were judged to be unrepresentative. Blinded test interpretation was not reported.

Miscellaneous other tests

One study evaluated three other miscellaneous tests: LDH, lymphocytes and lymphocyte proliferation.[223] Forty-four patients were included and TB prevalence was 18%. The reference standard used was culture plus histology. The study was prospective in design but the patient sample was again judged to be unrepresentative. Blinded test interpretation was not reported.

Results

Nucleic acid amplification tests

The single evaluation of a NAAT test included only seven patients and therefore no conclusions can be drawn regarding its accuracy.

Adenosine deaminase tests

The sensitivity and specificity pairs for the eight ADA evaluations were all clustered in the top

TABLE 62 *Peritoneal TB – summary of data sets identified*

	All	Ascites	Serum	Other
Total	14	12	1	1
NAAT tests	**1**	**1**	**0**	**0**
In-house	1	1	0	0
IS6110	1	1	0	0
Adenosine deaminase tests	**8**	**6**	**1**	**1** a/s ratio
ADA	8	6	1	1 a/s ratio
Cytokine tests	**2**	**2**	**0**	**0**
Interferon-γ	1	1	0	0
Total protein	1	1	0	0
Other miscellaneous tests	**3**	**3**	**0**	**0**
Lactate dehydrogenase (LDH)	1	1	0	0
Lymphocytes	2	2	0	0

a/s, ascites/serum.

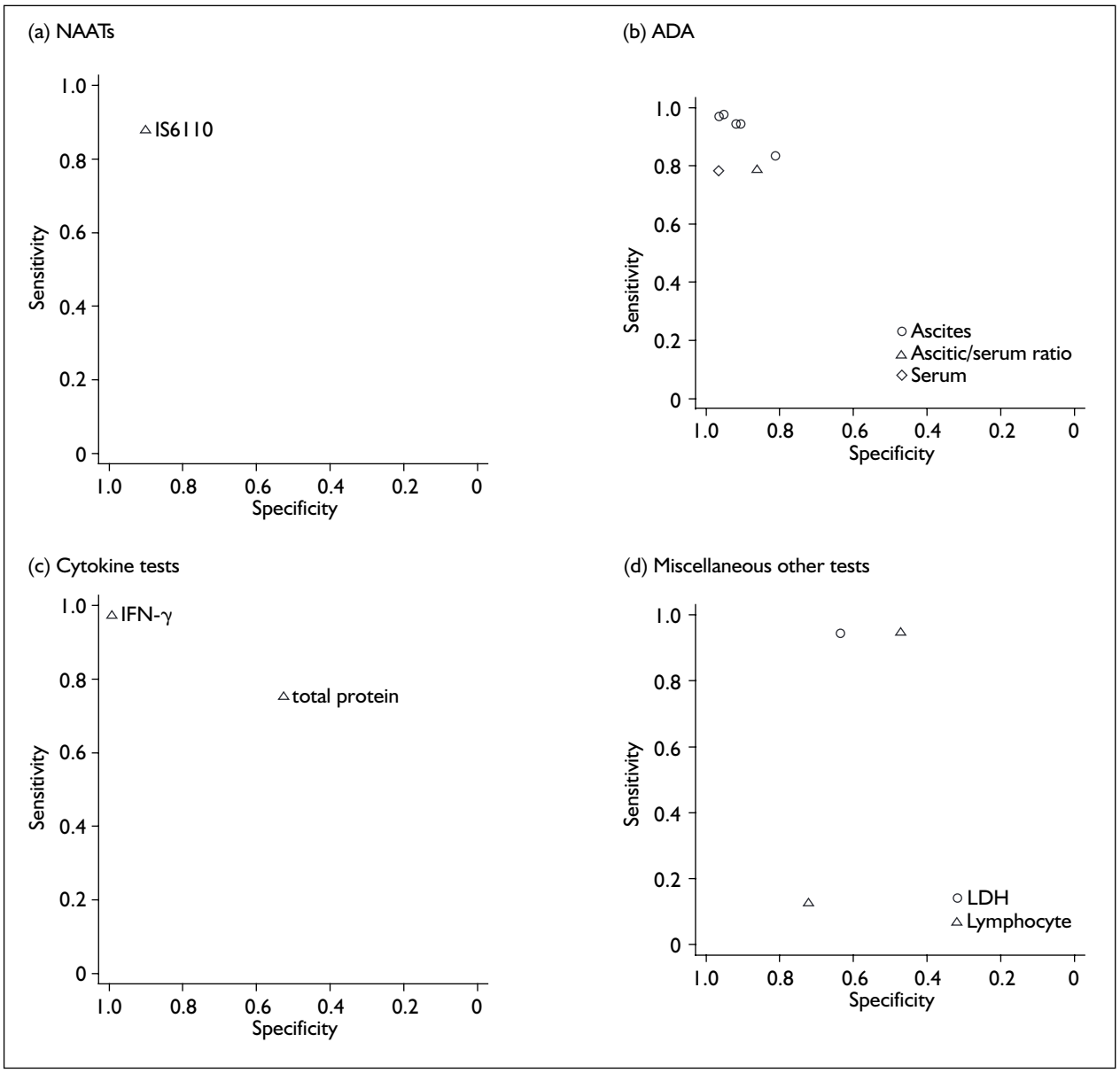


FIGURE 18 *Peritoneal TB – ROC plots*

TABLE 63 *Peritoneal TB: tests for heterogeneity of sensitivities and specificities*

		Sensitivity		Specificity	
Comparison	***n***	***Q statistic***	***p*-Value**	***Q statistic***	***p*-Value**
Adenosine deaminase tests	8	7	0.42	19	0.01
Cytokine tests	2	2.1	0.15	11.0	<0.01
Miscellaneous other tests	3	10	0.01	5	0.09

left-hand quadrant of the ROC plot (*Figure 18b*), although specificty estimates were still found to be statistically heterogeneous ($p = 0.01$) (*Table 63*). The test performed particularly well when applied to ascitic fluid samples (summary sensitivity 95.3% and specificity 93.2%) (*Table 64*). None of the data sets met all five design-related criteria; however, it is notable the single study that used blinded test interpretation reported much lower accuracy (sensitivity 83% and specificity 81%)[225] (Appendix 20).

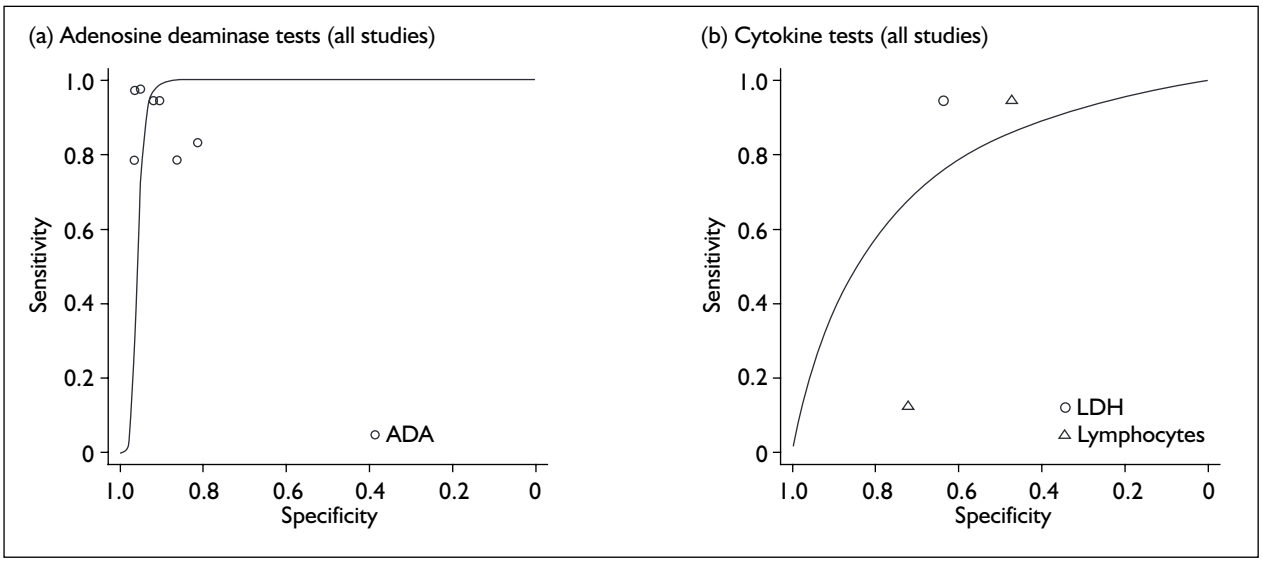


FIGURE 19 *Peritoneal TB – SROC curves*

Cytokine tests

Of the two cytokine test evaluations, interferon-γ was far superior at detecting TB infection, with both sensitivity and specificity of 100%[224] (*Figure 18b*). The study was both hospital based and used a combined reference standard, but did not use blinded test interpretation, nor were patients considered to be representative. Total protein was not found to be a good marker of TB infection (Appendix 20).

Miscellaneous other tests

Brant and co-workers[223] found LDH and lymphocyte proliferation to be highly sensitive for detecting TB infection but specificity was poor (Appendix 20 and *Figure 18d*).

Discussion: test accuracy in peritoneal TB

Of all the tests evaluated, there is some indication that ADA in ascitic fluid may be a good indicator of TB infection. However we were not able to determine test accuracy in studies using blinded test interpretation as this information was rarely reported. Other tests, including NAATs and interferon-γ, have not been evaluated in a sufficient number of quality studies to be able to make any comment on their potential use.

TABLE 64 *Peritoneal TB – SROC models*

Comparison	All studies					Studies meeting two design-related criteria				
	n[a]	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	*n*[a]	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
Adenosine deaminase tests										
ADA	8	184.00 (48.56 to 697.21)	0.36	92.9	93.2	5	273.81 (25.38 to 2954.11)	NA	92.4	94.5
By sample										
Ascitic fluid samples	6	276.90 (58.85 to 1302.90)	NA	95.3	93.2	3	550.34 (57.20 to 5295.42)	NA	96.4	95.3
Ascitic/serum samples	1	22.92 (4.37 to 120.10)	NA	78.6	86.2	1	22.92 (4.37 to 120.10)	NA	78.6	86.2
Serum only	1	102.67 (9.61 to 1096.34)	NA	78.6	96.6	1	102.67 (9.61 to 1096.34)	NA	78.6	96.6
Cytokine tests	2	124.91 (0.00 to >1,000,000)	NA	90.0	92.6	1	4653.00 (89.01 to 243,225.70)	NA	97.1	99.3
Miscellaneous tests	3	5.51 (0.02 to 1937.62)	NA	77.6	61.5					
LDH	1	29.59 (1.58 to 554.04)	NA	94.4	63.5	0				
Lymphocytes	2	2.38 (0.00 to >1,000000)	NA	60.9	60.4	0				

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

Chapter 11

Results: detection of pericardial tuberculosis infection

We identified seven test comparisons in pericardial TB (*Table 65*). ROC plots of each sensitivity and specificity pair for each group of tests are provided in *Figure 20(a–c)*.

Description and quality of included studies

The seven test evaluations were carried out in six studies; summary details per study are provided in Appendix 21.

TABLE 65 *Pericardial TB – summary of data sets identified*

	All
Total	7
NAAT tests	**2**
Commercial	1
Amplicor	1
In-house	1
IS6110	1
Adenosine deaminase tests	**4**
ADA	4
Cytokine tests	**1**
Interferon-γ	1

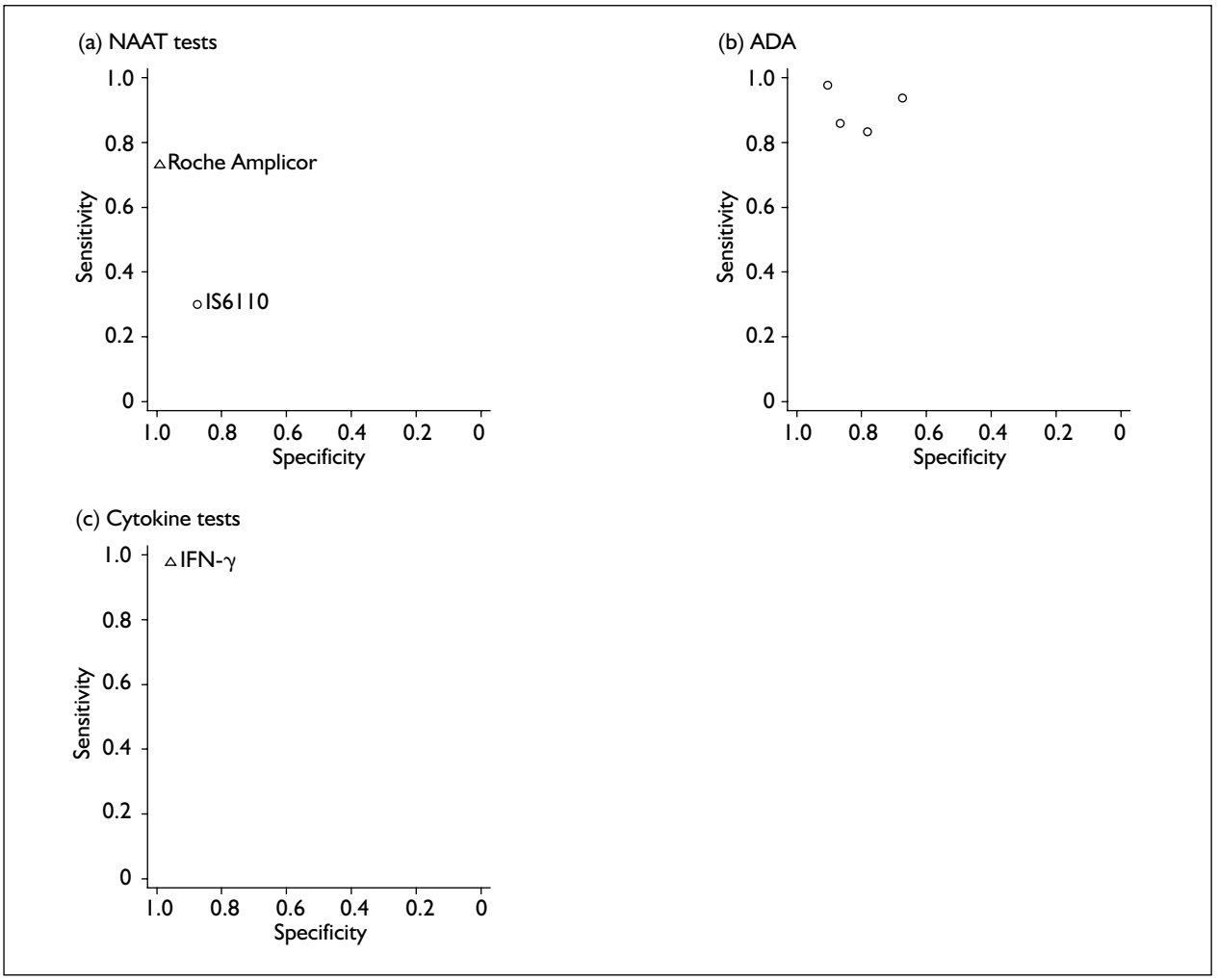


FIGURE 20 *Pericardial TB – ROC plots*

TABLE 66 *Pericardial TB – tests for heterogeneity of sensitivities and specificities*

		Sensitivity		Specificity	
Comparison	***n***	***Q* statistic**	***p*-Value**	***Q* statistic**	***p*-Value**
Adenosine deaminase tests	4	3	0.38	11	0.01

TABLE 67 *Pericardial TB – SROC models*

	All studies					Studies meeting two design-related criteria				
Comparison	***n*[a]**	**DOR (95% CI)[b]**	**Test of symmetry[c]**	**Se[d] (%)**	**Sp[d] (%)**	***n*[a]**	**DOR (95% CI)[b]**	**Test of symmetry[c]**	**Se[d] (%)**	**Sp[d] (%)**
NAAT tests	2	30.06 (0.00 to >1,000,000)	NA	51.9	96.5	1				
Adenosine deaminase tests										
ADA	4	53.85 (6.22 to 466.51)	NA	92.1	82.2	3	27.88 (10.33 to 75.23)	NA	88.5	78.4

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

Nucleic acid amplification tests

Two studies evaluated NAAT tests in pericardial fluid were included, one using the Amplicor test[226] and the other an in-house test using IS6110 target sequence.[222] The Amplicor study was large and of better quality than the IS6110 study (Appendix 21, Pericardial TB – study characteristics). It included 67 patients, used a combined reference standard and reported blinded index test interpretation. The IS6110 evaluation included only seven patients and prevalence was high at 57%.

Adenosine deaminase tests

Four evaluations of the ADA test were included. Three used a combined culture plus clinical diagnosis reference standard and two were prospective in design (Appendix 21). Two studies were judged to have included an unrepresentative patient sample.

Cytokine tests

One study assessed interferon-γ in pericardial fluid from 30 patients.[227] The study was prospective and used a combined reference standard but blinded test interpretation was not reported and prevalence was high at 63%.

Results

Nucleic acid amplification tests

The IS6110 test performed very poorly but only seven patients were tested. Specificity of the Amplicor test was 100% and sensitivity 75% (Appendix 21 and *Figure 20*).

Adenosine deaminase tests

The specificity estimates for the four ADA evaluations were found to be statistically heterogeneous ($p = 0.01$), but sensitivities were homogeneous (*Table 66*). The summary DOR from the SROC model was fairly high at 53.85, giving a summary sensitivity of 92% and a specificity of 82% (*Table 67* and *Figure 21b*). None of the data sets met all five design-related criteria; however, the two that reported blinded test interpretation had lower sensitivity estimates than the other two studies (*Table 66*).

Cytokine tests

The interferon-γ evaluation[227] reported 100% sensitivity and specificity; however, given the reservations regarding study quality noted above, not much weight can be put on this result.

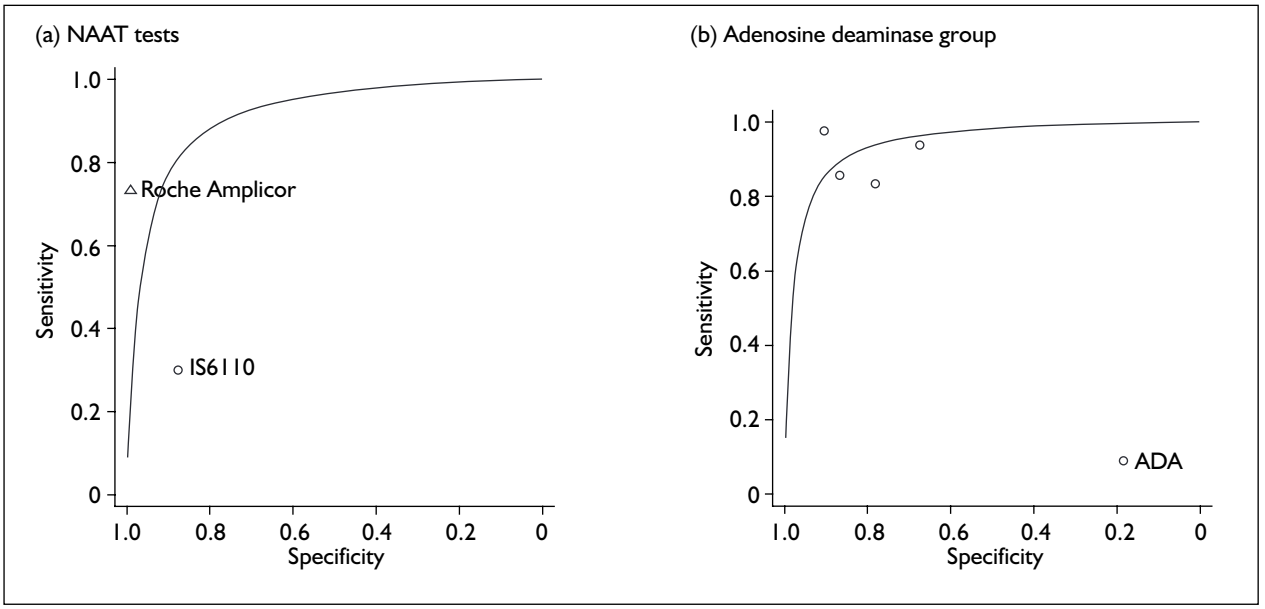


FIGURE 21 *Pericardial TB – SROC curves*

Discussion: test accuracy in pericardial TB

Of all the tests evaluated, there is some suggestion that ADA may be a good indicator of TB infection. However, when blinded index test interpretation was used, test accuracy was much lower. Further evaluation is needed before the test can be recommended for use. Other tests, NAAT tests and interferon-γ, have not been evaluated in a sufficient number of quality studies to make any comment on their potential use.

Chapter 12

Results: detection of genito-urinary tuberculosis infection

We identified 10 test comparisons in genito-urinary TB (*Table 68*). ROC plots of each sensitivity and specificity pair for each group of tests are provided in *Figure 22(a)* and *(b)*.

TABLE 68 *Genitourinary TB – summary of data sets identified*

	All	Urine	Serum
Total	10	9	1
NAAT tests	**9**	**9**	**0**
Commercial tests	5	5	
AMTD	3	3	
LCx	1	1	
Amplicis (Myco B)	1	1	
In-house tests	4	4	
16SrRNA	1	1	
IS6110	2	2	
MPB64	1	1	
Antigen tests	**1**	**0**	**1**
H37Rv	1	0	1

Description and quality of included studies

The 10 test evaluations were carried out in eight studies; a summary of key characteristics across all the NAAT tests is given in *Table 69*, with study characteristics for all test evaluations provided in Appendix 22.

Nucleic acid amplification tests

Of the nine NAAT test evaluations, five were of commercial tests and four of in-house tests (*Table 68*). The mean sample size across all data sets was 267 (SD 416), owing to one study[228] (providing two data sets) of 1000 patients; the remaining studies all had less than 100 patients each. TB prevalence was 30% (SD 24) (*Table 69*). Three of the nine data sets were from laboratory-based studies and one[229] used a combined culture plus clinical diagnosis reference standard. Two studies reported using blinded test interpretation; this information was not reported in the others. The study design could not be determined in any of the studies. The patient samples were judged to be representative in seven data sets.

Antigen tests

The single evaluation of an antigen test used the H37Rv antigen in 50 patients.[229] The sample was judged to be unrepresentative and binding and study design could not be determined. TB prevalence was 70%.

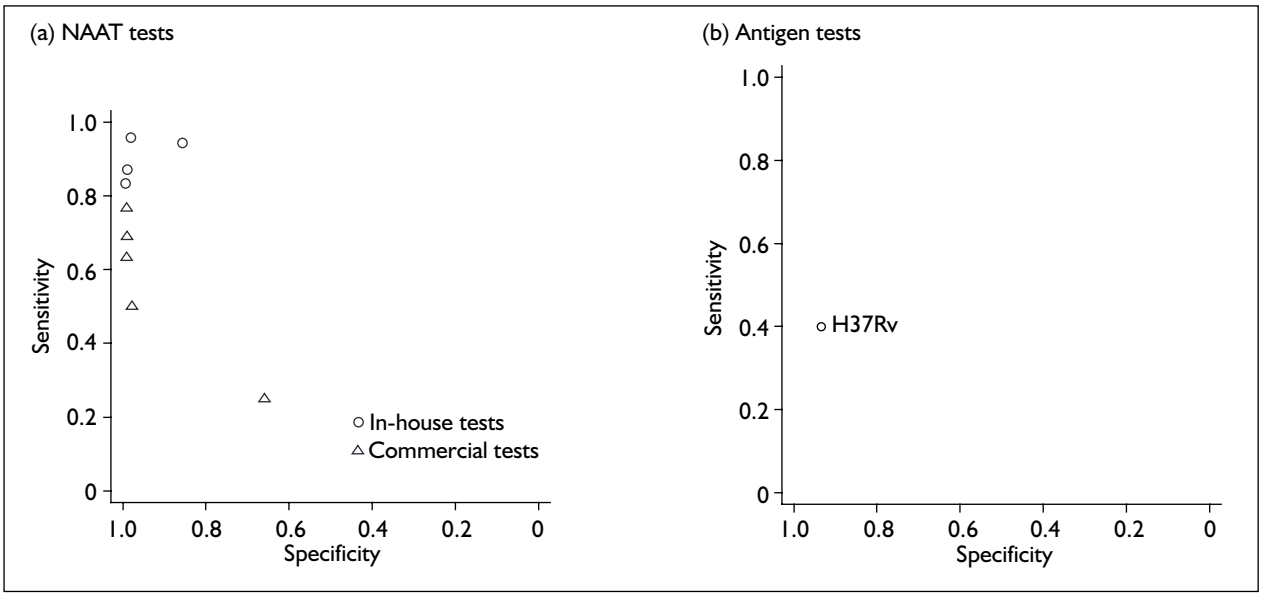


FIGURE 22 *Pericardial TB – SROC curves*

TABLE 69 *NAAT test evaluations in genitourinary TB – summary of key study characteristics*

	Respiratory samples
Total no. of studies	9
Mean sample size	267 (SD 416; range 22–1000)
Commercial	5
Mean sample size	55 (SD 23; range 22–73)
In-house	4
Mean sample size	531 (SD 541; range 42–1000)
Reference standard	
Culture + clinical diagnosis ± other	2 (22%)
Culture + anti-TB therapy	3 (33%)
Culture alone	4 (44%)
Clinical diagnosis alone	0
Disease prevalence (mean, SD, range)	30.4% (SD 24.4; range 2–83%)
Setting	
Hospital-based	6 (67%)
Laboratory-based	3 (33%)
Unknown	0
Sample type	
Urine	9 (100%)
Patients representative?	
Yes	7 (78%)
No	0
Unclear	2 (22%)
Study design prospective?	
Yes	0
No	0
Unclear	9 (100%)
Index test interpreted blinded?	
Yes	2 (22%)
No	0
Unclear	7 (78%)
Reference test interpreted blinded?	
Yes	2 (22%)
No	0
Unclear	7 (78%)

Results

Nucleic acid amplification tests

Almost all of the NAAT test comparisons had high specificity estimates (*Figure 22*) and more variable sensitivity estimates. Both statistical tests for heterogeneity were statistically significant ($p < 0.01$) (*Table 70*). The overall DOR was 89.05 (95% CI: 2.60 to 3048.87), but sensitivity was low at 76.9%. Test performance was much better in the in-house tests (DOR 484.60; 95% CI: 85.66 to 2741.60), giving an associated sensitivity of 91.3% and specificity of 97.9% (*Table 71*). The single study meeting all five design-related criteria[229] found an in-house NAAT using the MPB64 target sequence to have a sensitivity of 94% and a specificity 86%, but only 42 patients were included and prevalence was high (83%) (*Table 71*).

Antigen tests

The antigen test evaluation[230] reported low sensitivity (40%) and higher specificity (93%) using the H37Rv antigen (*Table 72*).

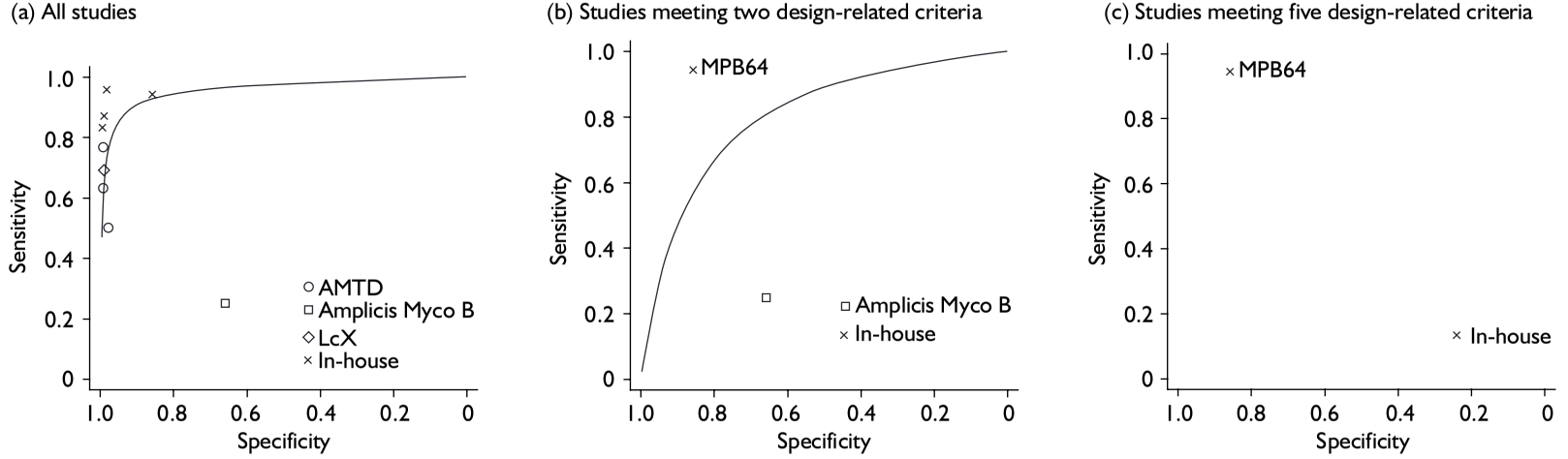
Discussion: test accuracy in genito-urinary TB

NAAT tests potentially have some role to play in the diagnosis of urinary TB infection. Most of the studies demonstrated high specificity, but few of these reported using blinded test interpretation. Further evaluation is needed before the test can be recommended for use. There is no evidence to support the use of antigen tests for detecting urinary TB.

TABLE 70 *NAAT tests in genito-urinary TB – tests for heterogeneity of sensitivities and specificities*

		Sensitivity		Specificity	
Comparison	***n***	***Q statistic***	***p*-Value**	***Q statistic***	***p*-Value**
All NAAT comparisons	9	56	<0.01	55	<0.01
Commercial tests	5	4	0.43	21	<0.01
AMTD	3	2	0.29	0	0.86
In-house tests	4	16	<0.01	6	0.09

TABLE 71 *NAAT tests in genito-urinary TB – SROC regression models and associated SROC curves*



	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se^d (%)	Sp^d (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se^d (%)	Sp^d (%)	n^a	DOR (95% CI)[b]	Test of symmetry[c]	Se^d (%)	Sp^d (%)
All tests	9	89.05 (2.60 to 3048.87)	0.71	76.9	97.8	2	7.99 (0.00 to >6,000,000)	NA	70.1	77.3	1	99.0 (7.71 to 1272.03)	NA	94.3	85.7
Commercial	5	0.01 (0.00 to 14.63)	0.03	57.1	97.6										
AMTD	3	153.0 (9.74 to 2420.14)	NA	64.1	97.0										
LcX	1	220.84 (11.73 to 4158.45)	NA	69.0	99.0										
Amplicis Myco B	1	0.64 (0.02 to 17.82)	NA	25.0	6.9	1	0.64 (0.02 to 17.82)	NA	25.0	67.0					
In-house	4	484.60 (85.66 to 2741.48)	NA	91.3	97.9	1	99.0 (7.71 to 1272.03)	NA	94.3	85.7	1	99.0 (7.71 to 1272.03)	NA	94.3	85.7
Subgroups meeting each design-related criterion															
Hospital-based	6	26.56 (0.12 to 6088.13)	0.55	95.4	99.7										
Combined reference test	2	7.99 (0.00 to >6,000,000)	NA	70.1	77.3										
Index test blinded	2	7.99 (0.00 to >100,000)	NA	70.1	77.3										
Reference test blinded	2	7.99 (0.00 to >100,000)	NA	70.1	77.3										
Patients representative.	7	294.73 (61.83 to 1404.93)	0.66	84.6	98.5										

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

TABLE 72 *Antigen tests in genitourinary TB – results*

Comparison	All studies					Studies meeting two design-related criteria				
	n[a]	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)	*n*[a]	DOR (95% CI)[b]	Test of symmetry[c]	Se[d] (%)	Sp[d] (%)
H37Rv	1	9.33 (1.10 to 79.21)	NA	40.0	93.3	0				

[a] Where <5 studies per subgroup, studies pooled assuming symmetric SROC curve (insufficient power to detect asymmetry); where only one study per subgroup, actual DOR and sensitivity and specificity (with zero cell correction) are reported.
[b] DOR where sensitivity = specificity.
[c] Test of symmetry indicates whether SROC curve is symmetric or not; if $p < 0.05$ SROC is asymmetric and indicates variation in threshold.
[d] Sensitivity and specificity from SROC curve, i.e. at mean of *s*.

Chapter 13

Results: detection of skeletal tuberculosis infection

Only one data set in skeletal TB was identified (see Appendix 23 and *Figure 23*).

Nucleic acid amplification tests

Van der Spoel van Dijk and colleagues[231] evaluated an in-house NAAT test using the IS6110 target sequence on spinal biopsy samples from 45 patients. TB prevalence was 29% and the patient sample was judged to be representative. The reference standard used was clinical diagnosis plus histology. Blinded test interpretation was not reported and the study design was not clear. Test sensitivity was 79% and specificity 89%.

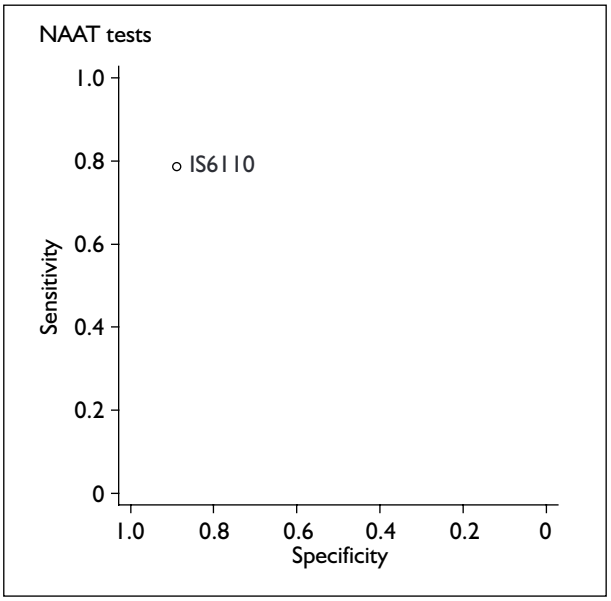


FIGURE 23 *Skeletal TB – ROC plots*

Chapter 14

Systematic review of fully automated liquid culture tests

Methods

The methods used for the fully automated liquid culture studies are as described in Chapter 3, apart from the following.

Inclusion criteria

Per specimen data

For the fully automated liquid culture tests, per specimen data were included, as none of the studies identified provided results on a per patient basis. Given the widespread adoption of this new and expensive group of tests, we judged it to be important to include a review of these studies despite this problem. There are possibilities of bias in estimates if the number of times an individual is tested and included in the data set is directly or indirectly related to their test results.

Diagnostic tests

Studies evaluating fully automated liquid culture techniques were included if a standard culture on liquid or solid media had also been performed in the same study or a comparison with the BACTEC 460 radiometric method had been performed.

Identification methods

A sample was determined to be a positive isolate if, in addition to at least one of the tested methods indicating growth (i.e. culture on solid media, BACTEC 460 or an automated system), some form of amplification or molecular probe test such as Accuprobe confirmed the isolate.

Study design and outcome measures

These studies only provide estimates of sensitivities. None of the studies identified verified all samples with a reference standard such as DNA hybridisation systems and biochemical and cultural tests to give 2 × 2 data – only those positive by at least one method were verified. A *post hoc* amendment was made to this section following screening of the FALC studies so that these studies could be included within the review. The performance of each method was determined by estimating the proportion of total isolates and of MTB isolates identified by each method (i.e. test sensitivity). Mean time to detection and contamination rates were also extracted.

Data synthesis

We compared culture methods by computing the relative false-negative rates between pairs of methods within each study and pooled them using a random effects model. As it is likely that studies that evaluated more tests missed fewer positive samples and reported lower sensitivities for each method, it is essential that comparisons are first made between the methods within studies and then pooled across studies. We opted to use relative false-negative rates rather than relative sensitivities to reduce heterogeneity and to avoid the bounding problem with relative proportions (if the sensitivity of solid culture is 80%, the highest ratio of sensitivities will be is bounded to a maximum value of 1.25).

For contamination rates, a similar meta-analysis was undertaken, comparing the relative proportion of contaminated specimens by each method.

Inadequate data were presented to be able to undertake a proper meta-analysis of detection times. Instead, the mean and range of times for each method and the mean time difference and range of time differences for each method are given. It is not possible to judge whether these differences are statistically significant.

Results

Description and quality of included studies

We identified 19 studies evaluating a FALC method plus at least one standard culture method. A matrix of the test comparisons performed is provided in *Table 73*, with details of the comparisons per study in *Table 74*. Study details are given in Appendix 24.

Tests

Four FALC methods were evaluated in the included studies (*Table 74*). The two main ones in

TABLE 73 *Matrix of test comparisons for detection of all isolates/MTB isolates*

Fully automated liquid culture methods	'Standard' culture methods		
	BACTEC 460	LJ	All solid
MGIT 960	4/4	4/4	7/7
MB/BacT	8/8	6/6	8/9
BACTEC 9000	0/1	1/1	1/1
ESP II	2/2	1/1	3/3

TABLE 74 *Fully automated liquid culture tests – summary of data sets identified*

Study	Fully automated liquid culture methods				'Standard' culture methods		
	MGIT 960	MB/BacT	BACTEC 9000	ESP II	BACTEC 460	LJ	Other solid
Alcaide, 2000[120]	✓	✓			✓	✓	
Benjamin, 1998[457]		✓			✓		
Brunello, 1999[458]		✓			✓	✓	
Gil-Setas, 2004[76]		✓					✓
Hanna, 1999[77]	✓				✓		✓
Harris, 2000[460]		✓			✓		✓
Idigoras, 2000[232]	✓						✓
Kanchana, 2000[455]	✓				✓	✓	
Lu, 2002[456]	✓					✓	
Palacios, 1999[461]		✓				✓	
Piersimoni, 2001[462,463]		✓			✓	✓	
Roggenkamp, 1999[464]		✓			✓	✓	
Rohner, 1997[465]		✓			✓		✓
Somoskovi, 2000[466]		✓			✓	✓	
Tortoli, 1998[468]				✓	✓	✓	
Tortoli, 1999[89]	✓				✓	✓	
Van Griethuysen, 1996[467]			✓			✓	
Williams-Bouyer, 2000[469]	✓			✓			✓
Woods, 1997[470]				✓	✓		✓
Total no. of studies	7	10	1	3	13	11	7

current use are the BACTEC MGIT 960 (six studies) and the MB/BacT method (10 studies). The BACTEC 9000 (one study) and ESP II (three studies) are either early 'test-bed' systems or are not in widespread use.

The FALC methods have variously been compared with the radiometric liquid culture BACTEC 460 or with culture on solid media, predominantly LJ but also Middlebrook 7H11 or non-specified egg-based media (*Table 73*).

Reference tests

None of the studies used an independent 'reference test' *per se*; instead, the number of isolates identified by each individual culture method was compared with the total number of isolates identified by all methods, both fully automated and more standard methods (i.e. the reference standard was the maximum number of isolates cultured).

Identification of isolates

The cultured isolates were identified as *M. TB* or other mycobacteria primarily using the Accuprobe test and biochemical and cultural tests. In eight studies Accuprobe was performed only on AFB-positive samples. One study[120] used PCR restriction fragment length polymorphism analysis on all isolates and one[76] used the AMTD test.

Sample details

All studies bar one[232] included mixed respiratory and non-respiratory samples. One study did not report the proportion of samples that were respiratory, but for the remainder, the mean proportion that were respiratory was 72% (SD 11%, range 56–100%). Thirteen studies reported

TABLE 75 *Comparisons of false-negative rates with solid culture and BACTEC 460*

Method	Detection of any isolate				Detection of MTB isolates			
	Comparison with all solid culture		Comparison with BACTEC 460		Comparison with all solid culture		Comparison with BACTEC 460	
	n	RFN (95% CI)	*n*	RFN (95% CI)	*n*	RFN (95% CI)	*n*	RFN (95% CI)
BACTEC 9000	1	0.20 (0.09 to 0.47)	0	–	1	0.10 (0.02 to 0.40)	0	–
ESP II	3	0.57 (0.40 to 0.82)	2	1.15 (0.42 to 3.19)	3	1.20 (0.67 to 2.17)	2	3.09 (0.75 to 12.68)
MB/BacT	8	0.52 (0.32 to 0.86)	8	1.57 (0.98 to 2.52)	9	0.40 (0.25 to 0.64)	8	1.50 (0.72 to 3.12)
MGIT 960	7	0.50 (0.33 to 0.76)	4	0.88 (0.57 to 1.35)	7	0.53 (0.33 to 0.85)	4	0.74 (0.40 to 1.36)

RFN, relative false-negative rate.

TABLE 76 *Comparison of time to detection with solid media*

Method		Time to detection (days)		
	n	Automated method: mean [min., max.]	Solid media: mean [min., max.]	Difference: mean [min., max.]
All isolates				
BACTEC 9000	1	17.6 [–, –]	29.4 [–, –]	–11.8 [–, –]
ESP II	3	16.2 [13.1, 18.1]	21.6 [17.8, 27.8]	–5.4 [–9.7, –2.3]
MB/BacT	9	14.3 [11.7, 17.5]	23.3 [15.6, 26.8]	–10.0 [–2.2, –15.1]
MGIT 960	7	12.5 [11.1, 13.3]	22.1 [19.2, 15.7]	–8.9 [–6.1, –12.4]
TB isolates				
BACTEC 9000	0	–	–	–
ESP II	3	17.9 [15.5, 19.1]	22.0 [18.3, 28.6]	–4.1 [–0.5, –9.5]
MB/BacT	9	16.2 [13.3, 18.1]	23.9 [18.9, 35.8]	–7.7 [–3.1, –21.5]
MGIT 960	7	13.1 [10.5, 14.4]	20.2 [13.0, 25.2]	–7.4 [–0.3, –14.7]

the proportion of all isolates grown that were AFB positive: mean 51% (SD 17, range 25–88%). Eleven studies reported the proportion of samples growing *M. TB* that were AFB positive: mean 62% (SD 14, range 35–84%). The mean proportion of total samples growing non-tuberculous mycobacteria across the 19 studies was 3.3% (SD 1.9, range 1–7%).

Results

False-negative rates

Each of the four automated methods (BACTEC 9000, ESP II, MB/BacT and MGIT 960) was compared with all solid culture, LJ culture and BACTEC 460 to evaluate their relative sensitivities. Comparisons were first made for the detection of any isolate and second for MTB isolates only. Results are shown in *Table 75* and *Figure 24* for comparisons with solid culture and BACTEC 460. Results against LJ alone were similar to those against all solid culture and are not shown.

All methods were significantly better at detecting isolates than solid culture or LJ culture, roughly halving the number of missed positive results. The comparison of the BACTEC 9000 system with culture was based on only one study. The BACTEC 9000, MB/BacT and MGIT methods also significantly reduced the number of missed MTB isolates, and the ESP II method was not significantly better than solid culture.

No method was significantly better than the BACTEC 460 method for detection of either any isolate or MTB isolates only. No data were available for this comparison for the BACTEC 9000 system. Although not statistically significant, the direction of the estimates suggests that the ESP II and MB/BacT methods may be less accurate than the BACTEC 460 method.

Time to detection

Mean time to detection of any isolate and MTB isolates for the four systems are given in *Table 76* compared to solid culture, and in *Table 77* compared with the BACTEC 460. All studies reported faster detection times with the automated method compared with solid culture, with average

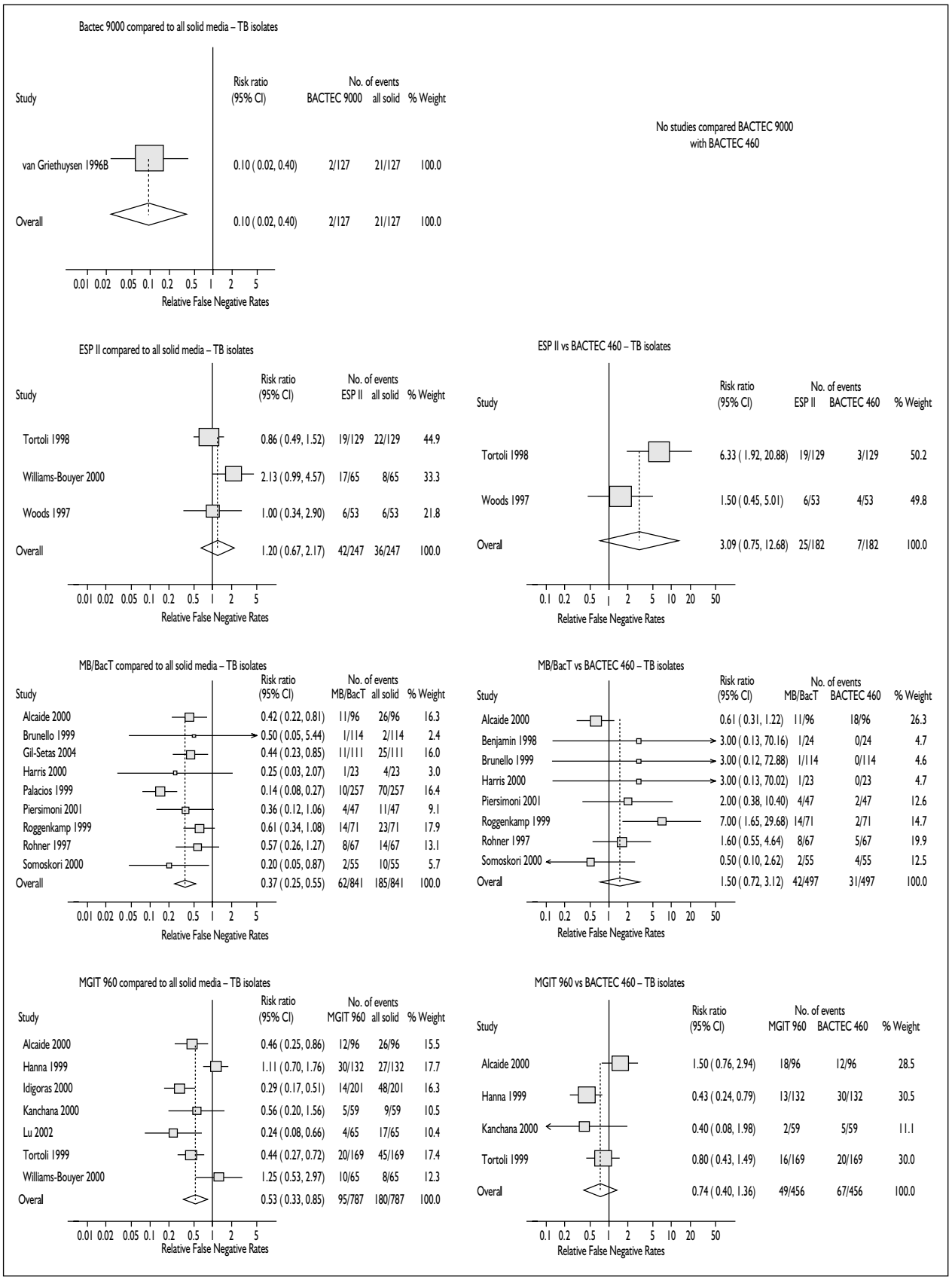


FIGURE 24 *Forest plots for relative false-negative rates*

TABLE 77 *Comparison of time to detection with BACTEC 460*

Method		Time to detection (days)		
	n	**Automated method: mean [min., max.]**	**BACTEC 460: mean [min., max.]**	**Difference: mean [min., max.]**
All isolates				
BACTEC 9000	0	–	–	–
ESP II	2	15.8 [13.1, 18.1]	16.2 [14.4, 17.8]	–0.4 [-1.3, 0.3]
MB/BacT	8	14.3 [11.7, 17.5]	11.9 [9.2, 14.3]	2.3 [0.5, 4.2]
MGIT 960	4	12.2 [11.1, 13.3]	12.8 [11.3, 14.8]	–0.6 [–1.5, 0.6]
TB isolates				
BACTEC 9000	0	–	–	–
ESP II	2	17.4 [15.5, 19.1]	17.7 [16.6, 18.6]	–0.2 [–1.1, 0.5]
MB/BacT	8	16.0 [13.3, 18.1]	13.3 [9.6, 16.6]	2.7 [–2.3, 4.2]
MGIT 960	4	13.4 [10.5, 14.4]	14.1 [11.8, 16.0]	–0.7 [–5.5, 1.2]

TABLE 78 *Comparison of contamination rates with solid media*

Method		Contamination rates (%)		
	n	**Automated method: mean [min., max.]**	**Solid media: mean [min., max.]**	**Relative rate (95% CI)**
BACTEC 9000	1	6 [–, –]	6.5 [–, –]	0.92 (0.68 to 1.26)
ESP II	3	12.4 [7.9, 18.9]	7.4 [0.8, 11.0]	2.44 (1.05 to 5.69)
MB/BacT	7	7.3 [3.0, 10.0]	4.7 [1.2, 6.7]	1.34 (0.93 to 1.92)
MGIT 960	5	11.5 [3.3, 17.1]	8.8 [2.1, 17.1]	1.36 (0.76 to 2.42)

TABLE 79 *Comparison of contamination rates with BACTEC 460*

Method		Contamination rates (%)		
	n	**Automated method: mean [min., max.]**	**BACTEC 460: mean [min., max.]**	**Relative rate (95% CI)**
BACTEC 9000	0	–	–	–
ESP II	2	8.2 [7.9, 8.6]	4.0 [4.0, 4.0]	2.05 (1.74 to 2.42)
MB/BacT	8	5.3 [3.0, 7.0]	2.9 [1.6, 4.9]	1.79 (1.41 to 2.27)
MGIT 960	4	8.1 [3.3, 10.0]	4.0 [1.6, 4.9]	2.10 (1.44 to 3.05)

reductions of between 2 and 15 days for any isolates and between 1 and 21 days for detected MTB. As the studies poorly reported statistical analyses of time to detection, it is not possible to investigate formally whether these differences are statistically significant, or to make comparisons between the four methods. Again, only one study reported detection times for the BACTEC 9000 system, and then only for all isolates and not for MTB isolates.

Comparison of detection times with the BACTEC 460 system generally showed comparable times, some studies suggesting that the BACTEC 460 was faster and others reporting that it took longer, but the magnitude of the differences was small.

Contamination rates

The relative contamination rates are presented in *Table 78* for comparison with solid culture and in *Table 79* for comparison with the BACTEC 460 system. There was no significant difference between the BACTEC 9000, MB/BacT and MGIT 960 methods and solid culture media in contamination rates, but the ESP II system more than doubled the number of contaminated samples.

In comparison with the BACTEC 460 method, contamination rates with ESP II, MB/BacT and MGIT 960 were all significantly higher, with doublings of the number of contaminated samples. No data were available to compare the BACTEC 460 and BACTEC 9000 methods.

Discussion

Fully automated liquid culture methods are of value in terms of their speed and their precision for the detection of isolates compared to solid media. Although the BACTEC 460 radiometric method also has the same benefits, it is radiometric and therefore requires disposal of radioactive waste and also requires more staff time than the fully automated methods.

MGIT 960 and MB/BacT have demonstrated their superiority to solid culture in terms of reducing false-negative rates and detection times for MTB isolates. They have higher contamination rates than the BACTEC 460 radiometric method, but appear comparable in terms of detection times and detection rates. Although the ESP II method showed reductions in time to detection, it also had significantly higher contamination rates and showed no increase in accuracy over solid culture. There were insufficient data to evaluate the BACTEC 9000, with no data available that reported results separately for MTB isolates. Data on contamination rates may be somewhat outdated, as recent developments such as the addition of antibiotics during the process may have reduced them. We cannot comment from the studies we located on the incremental benefit of using more than one method (such as a rapid method and solid culture).

Chapter 15

Systematic review of tests for detection of latent tuberculosis infection

Methods

A systematic review was conducted to evaluate if a particular type of interferon-γ assay is superior in the diagnosis of latent *M. TB* infection. For this review, it was hypothesised that the association between the interferon-γ assay results and tuberculosis exposure will be stronger than that between TST results and exposure. The review also explored whether TST was more strongly associated with BCG vaccination than interferon-γ assays and if assays were less likely to cause false-negative results in HIV co-infected individuals than TST.

Identification of studies

The electronic searches targeted all diagnostic evaluations of interferon-γ assays in TB. Studies were identified from various sources. MEDLINE (1966–March 2004), EMBASE (1980–March 2004) and CAB Abstracts (1973–March 2004) were searched electronically. The search term combination evaluated the following concept: interferon-γ assays AND *M. TB* AND adapted sensitive diagnostic search filter[233–235] (Appendix 25). For completeness, individual experts with an interest in this field and the authors of relevant studies were contacted to uncover grey literature. Reference lists of known reviews and primary articles and related websites were also checked to identify cited articles not captured by electronic searches and personal contacts.

Study selection and data extraction procedures

The study selection criteria were (i) testing for LTBI, (ii) comparison between TST and interferon-γ assays based on ESAT-6 and CFP-10 including ELISPOT and whole blood ELISA and (iii) information on TB exposure or BCG vaccination or HIV status. Studies were selected in a two-stage process. First, the electronic searches were scrutinised and full texts of all citations that were likely to meet the predefined selection criteria were obtained. Second, final inclusion or exclusion decisions were made on examination of these texts. In cases of duplicate publication, the most recent and complete versions were selected. There were no language restrictions.

Information was extracted from each selected article on study characteristics, quality and results. Study characteristics consisted of patients' classifications and test characteristics including TST and interferon-γ assays based on PPD- or RD1-based antigen assays (ESAT-6, CFP-10). Data were used to construct contingency tables of test results according to TB exposure, BCG vaccination status and HIV status. When there were several categories of TB exposure, the information on a gradient or 'dose–response' was extracted from the reported analyses. In order to make comparisons between studies in the review, data were collapsed into two categories, one of higher exposure and another of lower exposure, to generate the 2×2 tables.

Assessment of methodological features

All publications meeting the selection criteria were assessed for their methodological features. Methodological quality was defined as the confidence that the study design, conduct and analysis minimised bias[236] in the estimation of association of the test with disease exposure. Using currently available checklists and texts on evaluation of observational studies,[236–244] a study was considered to be of good quality if it used the following features: a prospective design, consecutive enrolment of participants, adequate test description (to allow replication by others), blinding of the interferon-γ assay from TST result and vice versa and detailed assessment for exposure status. However, to quantify precisely exposure to active TB is often very difficult. Studies that were able to define exposure in an outbreak setting accurately were considered better for this review than those using comparison of household contacts with healthy controls, because outbreak investigations confirm the diagnosis among cases, capture all contacts and evaluate closeness and length of time spent with cases among contacts in detail. Studies that satisfied more than two-thirds of the quality features were considered to be of high quality.

Data synthesis

The main analysis examined whether interferon-γ assays were more strongly associated with high versus low TB exposure than TST. Data were synthesised separately for studies on various types of assays (PPD-based assays and those based on ESAT-6 or CFP-10) and the corresponding findings of TST. ORs were calculated for the association between test results and exposures from each study along with their 95% CIs using Review Manager 4.2 software.[245] Where 2 × 2 tables contained zero cells, 0.5 was added to allow the calculations. Plots of ORs and their CIs were used to examine heterogeneity, the differences between test performances among studies.

The primary analysis examined whether interferon-γ assays or TST were more strongly associated with high versus low TB exposure. Within each study, the OR value for one test was divided by that for another to produce a ratio of odds ratios (ROR). When comparing the interferon-γ assay with TST, an ROR value >1 meant that the assay was more strongly associated with TB exposure than TST, whereas an ROR value <1 meant that the assay was less strongly associated with TB exposure than TST. Both OR and ROR, in this context, reflect test performance and provide an approach to evaluating tests in the absence of a gold standard.[246] OR is a function of test sensitivity and specificity and increases as one or both of these measures increase. Statistically OR = [sensitivity/(1 – specificity)] / [(1 – sensitivity)/specificity]. When ROR > 1, the OR value for interferon-γ assay will be higher than that for TST, which means that either sensitivity or specificity or both associated with interferon-γ assay will be higher than those associated with TST.

We used results from individual studies to generate pooled RORs (meta-analysis) for comparing RD1-specific antigen-based assays with PPD-based assays, RD1-specific antigen-based assays with TST and PPD-based assays with TST. The log(ROR)s weighted according to the inverse of their variances were used to produce summary log(ROR)s and the output was exponentiated. The variance of log(ROR)s was generated using the equation[247]

$$\mathrm{Var}[\ln(\mathrm{ROR})] = \mathrm{Var}[\ln(\mathrm{OR}_1)] + \mathrm{Var}[\ln(\mathrm{OR}_2)] - 2r\sqrt{\mathrm{Var}[\ln(\mathrm{OR}_1)]\mathrm{Var}[\ln(\mathrm{OR}_2)]}$$

which took into account the correlation between tests, r. We assumed $r = 0.5$, which allowed us to produce conservative estimates. The square root of the variance gave the standard error, from which the 95% CIs were calculated for each ROR value. We used a random effects model for pooling results considering the heterogeneity.

A number of prespecified secondary analyses were performed. These included an assessment of whether the observed association varied according to study quality, test type, TB prevalence, BCG vaccination and immunosuppression (HIV infection). The last two analyses explored whether interferon-γ assays produced more accurate results than did TST in latent *M. TB* infection among patients with BCG vaccination and HIV infection. Using 2 × 2 tables of test results and BCG or HIV status, ORs were calculated for the tests. To measure the degree to which false results were more likely with one or another test, RORs were produced followed by meta-analysis in the same manner as that explained above. In the BCG vaccination analysis, when comparing interferon-γ assay with TST, ROR < 1 meant that false-positive results were less likely with the assay compared with TST. Conversely, in the HIV infection analysis, when comparing interferon-γ assay with TST, ROR > 1 meant that false-negative results were less likely with the assay compared with TST.

The findings of these analyses were used to determine differences in performance characteristics between interferon-γ assay types (assays based on PPD- and RD1-specific antigens). To aid in interpretation, the findings of the subgroups of highest quality studies included in the review were examined along with overall results.

Results

Selection of studies

There were 627 citations in the electronic searches excluding duplicates (*Figure 25*). Of the 168 citations in CAB abstracts, 115 were duplicates of MEDLINE and EMBASE searches. Scanning through the titles and abstracts (where available) of the electronic searches, 63 citations were considered potentially relevant. Their full texts were obtained and scrutinised. Many papers did not meet several of the study selection criteria. The bibliography of a recent systematic review[52] revealed six potentially relevant references,[248–253] which had to be excluded as they did not meet our inclusion criteria. A list of excluded studies with reasons is available.[254] One relevant citation[177] was identified by contact with authors. In total, 13 studies[153–155,158,166,177,178,255–260] were included in the review.

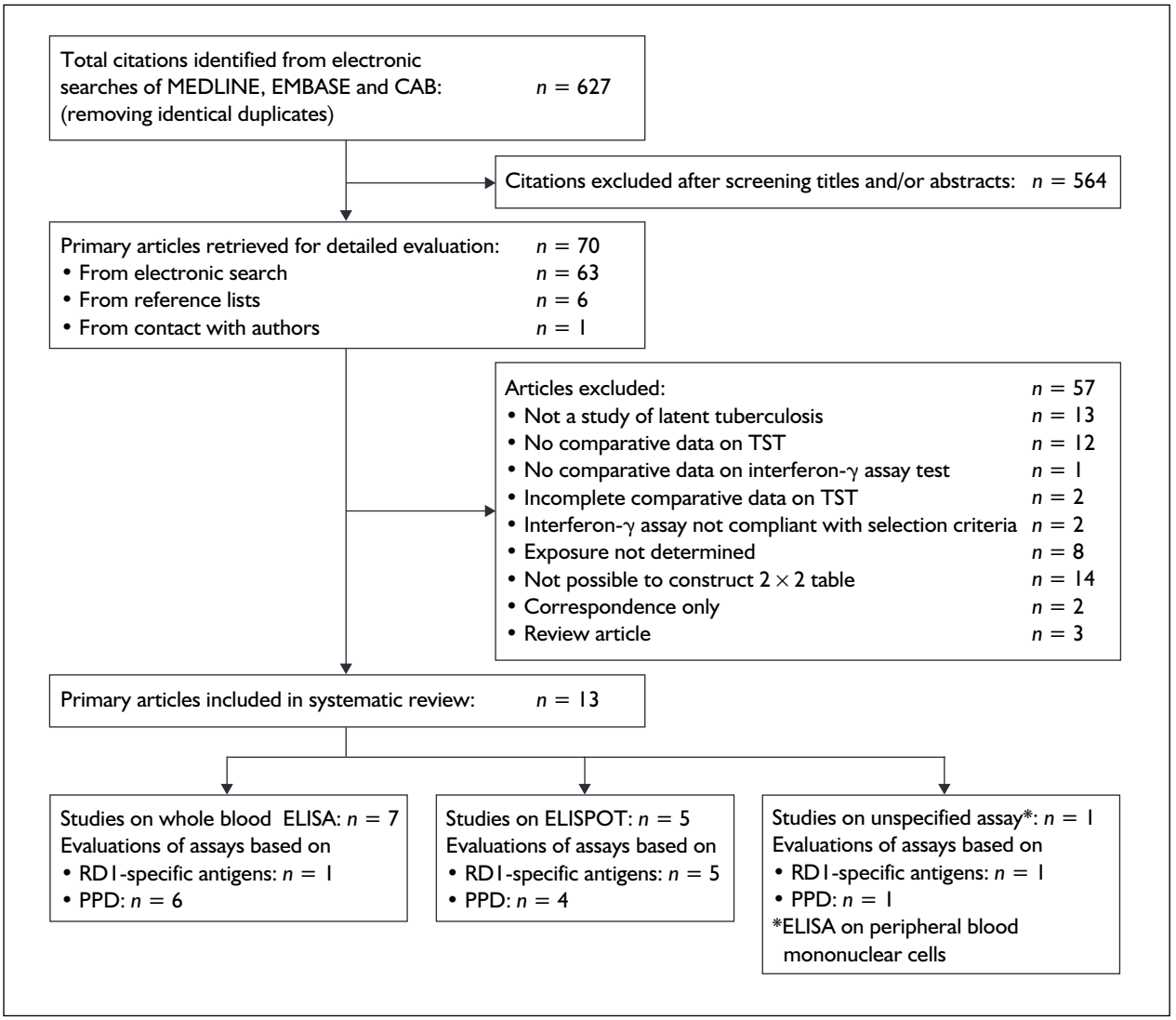


FIGURE 25 *Study selection process for systematic review of interferon-γ assays for latent tuberculosis*

Characteristics and quality of studies

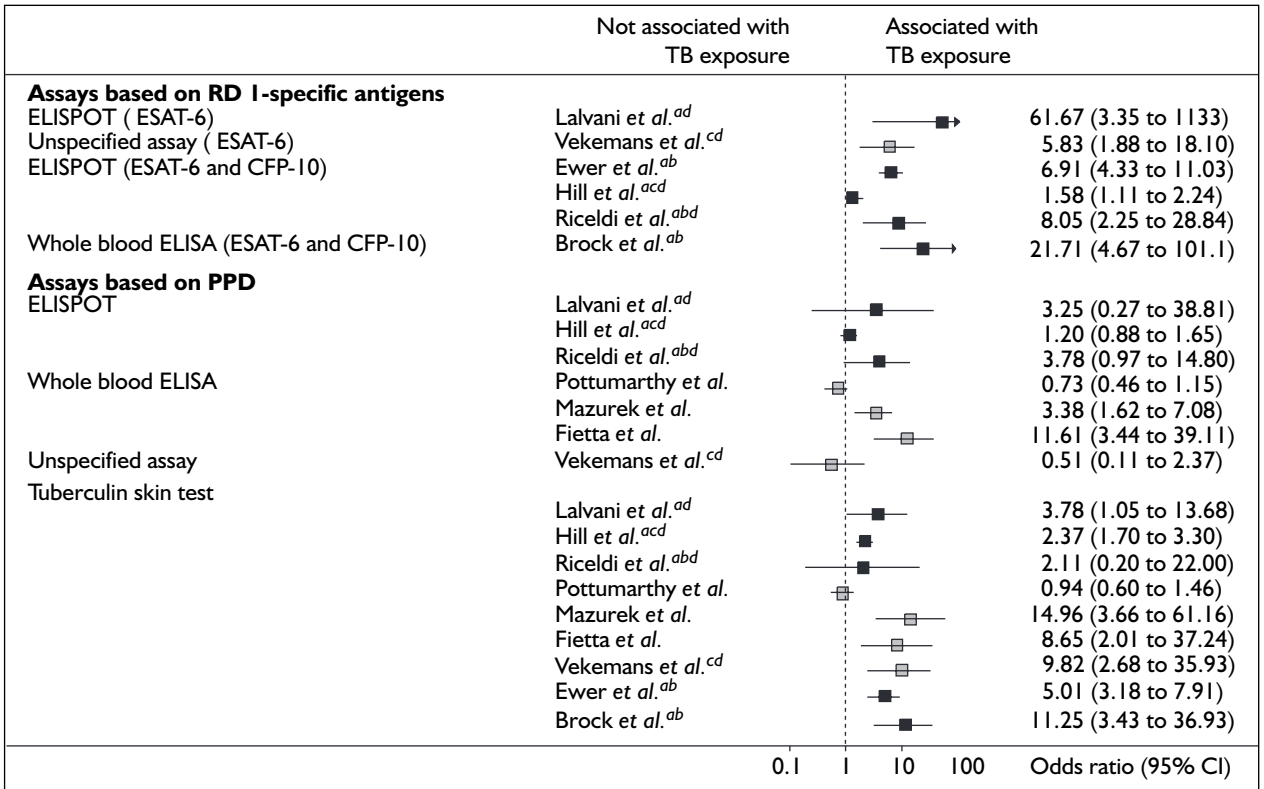
Five[153,166,177,178,258] studies used assays based on ELISPOT (n = 1461), seven[154,155,158,255–257,259] used whole blood interferon-γ ELISA (n = 2516) and one[260] used an unspecified interferon-γ assay [peripheral blood mononuclear cells (PBMC) ELISA] (n = 58), including a total of 4035 subjects (*Figure 25*, Appendix 27). Four studies provided no information on TB exposure.[153,255,257,259] Useful information to allow planned secondary analyses on BCG vaccination and HIV status was available in five[154,158,166,177,178] and four studies,[153,255,257,259] respectively. The performance of interferon-γ assays had been examined comparing test results among groups with high and low TB exposure in nine studies.[154,155,158,166,177,178,256,258,260] Four of the studies[153,255,258,260] (n = 1578) were carried out in high TB prevalence countries.

The methodological quality of studies was variable (Appendix 28). Although most studies seem to be prospective in design, only two recruited participants consecutively. Studies generally described the tests and their thresholds in sufficient detail. However, it was often not clear if staff were blinded to test results. Some studies reported blinding to either TST or interferon-γ assay, but not both. Ascertainment of TB exposure was adequate in only five of the studies.[166,177,178,256,258] These included three studies based on outbreak investigation.[177,178,256] In another study,[166] part of the sample included patients from an outbreak investigation.

Performance of the tests

The association of tests with TB exposure

The performance of interferon-γ assays was examined, comparing test results among groups



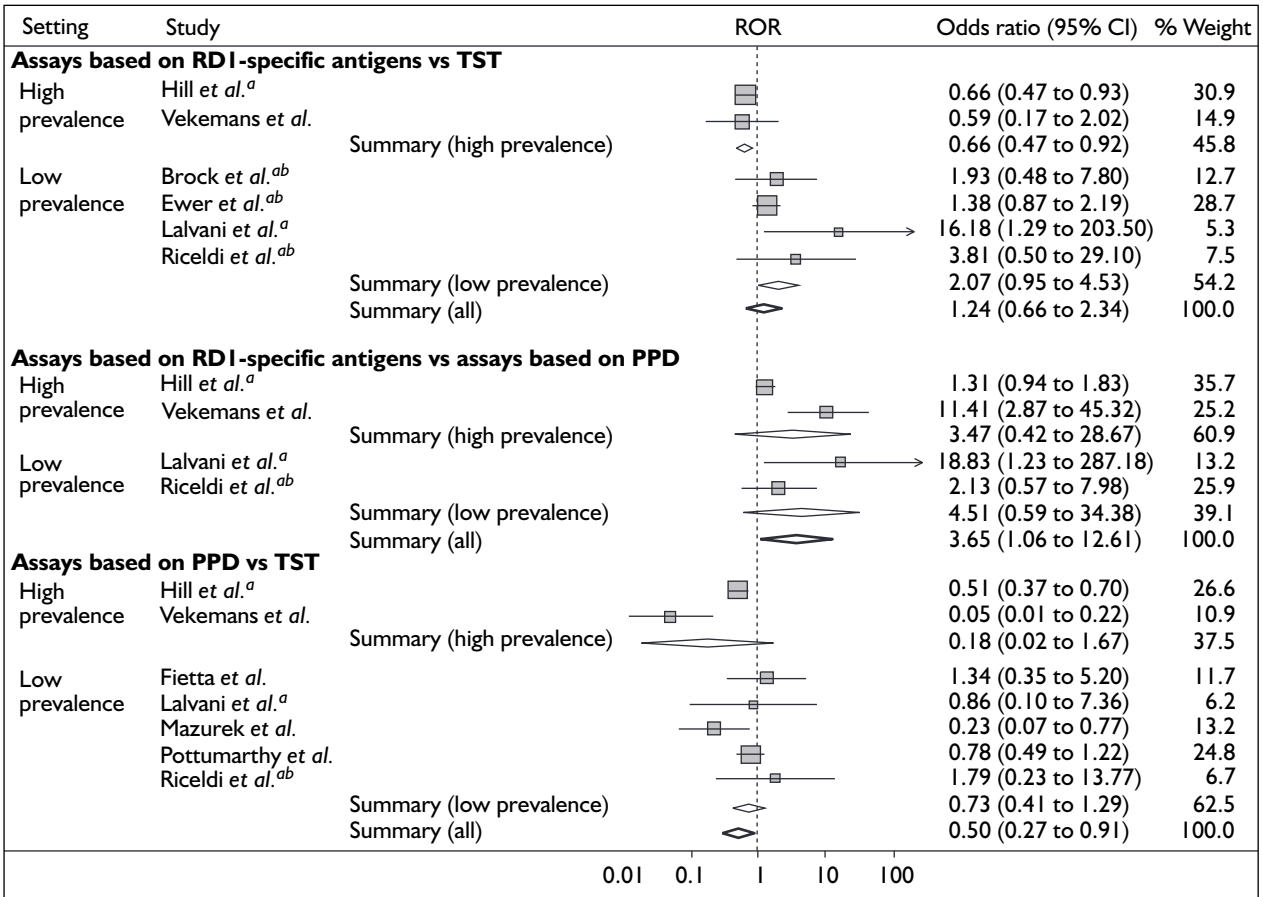
CFP, culture filtrate protein; ESAT, early secretory antigen target; PPD, purified protein derivative.
ORs and CIs generated using RevMan 4.2 software.[245]
Solid black squares represent point estimates of high-quality studies; grey squares indicate studies with moderate quality.
[a] Studies with adequate ascertainment of exposure.
[b] Studies investigating an outbreak.
[c] Studies in high TB prevalence country.
[d] RD1 specific antigen-based assays more strongly associated with TB exposure than PPD-based assays in direct comparison.

FIGURE 26 *Association of interferon-γ assays and TSTs with exposure to tuberculosis*

with high and low TB exposure in nine studies.[154,155,158,166,177,178,256,258,260] Three of the studies[177,178,256] assessed TB exposure to an index case in an outbreak investigation, one[166] partially in an outbreak investigation, two[166,258] exposure among contacts of active TB cases and four[154,155,158,260] compared high- with low-risk community groups. Some studies contained information on more than one type of assay. There were seven comparisons of interferon-γ assay based on PPD, three[166,177,258] using ELISPOT, three[154,155,158] using whole blood interferon-γ ELISA , and one[260] using an unspecified interferon-γ assay, with TST. There were two comparisons of interferon-γ assay based on ESAT-6, one[166] using ELISPOT, and one[260] using an unspecified interferon-γ assay, with TST. There were four comparisons of interferon-γ assay based on ESAT-6 and CFP-10 antigens, three[177,178,258] using ELISPOT and one[256] using whole blood ELISA, with TST.

Figure 26 shows that RD1-specific antigen-based assays had an association with exposure in 6/6 studies, PPD-based assays had an association in 2/7 studies and TST had an association in 7/9 studies, with 95% CI of OR excluding 1.0. Four studies[166,177,178,258] provided detailed information on the gradient of 'dose–response' relating degree of TB exposure to test results using tests for linear trends. Of these, two[178,258] studies showed a statistically significant relationship of exposure with TST, two[177,258] with PPD-based assays and all four[166,177,178,258] with RD1-specific antigen-based assays. Comparison of these gradients in one study[178] conducted in a low prevalence setting showed that an RD1-specific antigen-based assay was statistically significantly better than TST.

As shown in *Figure 27*, RD1-specific antigen-based assays were as strongly associated with TB exposure as TST overall (ROR 1.24; 95% CI: 0.66 to 2.34, $p = 0.49$), a finding that was based on a heterogeneous meta-analysis of studies from high



PPD, purified protein derivative.
See Appendix 29 for details for numerical data and *Figure 26* for the underlying ORs.
ROR provides comparison between tests: values >1 indicate association with exposure to tuberculosis.
[a] Studies with adequate ascertainment of exposure.
[b] Studies investigating an outbreak.

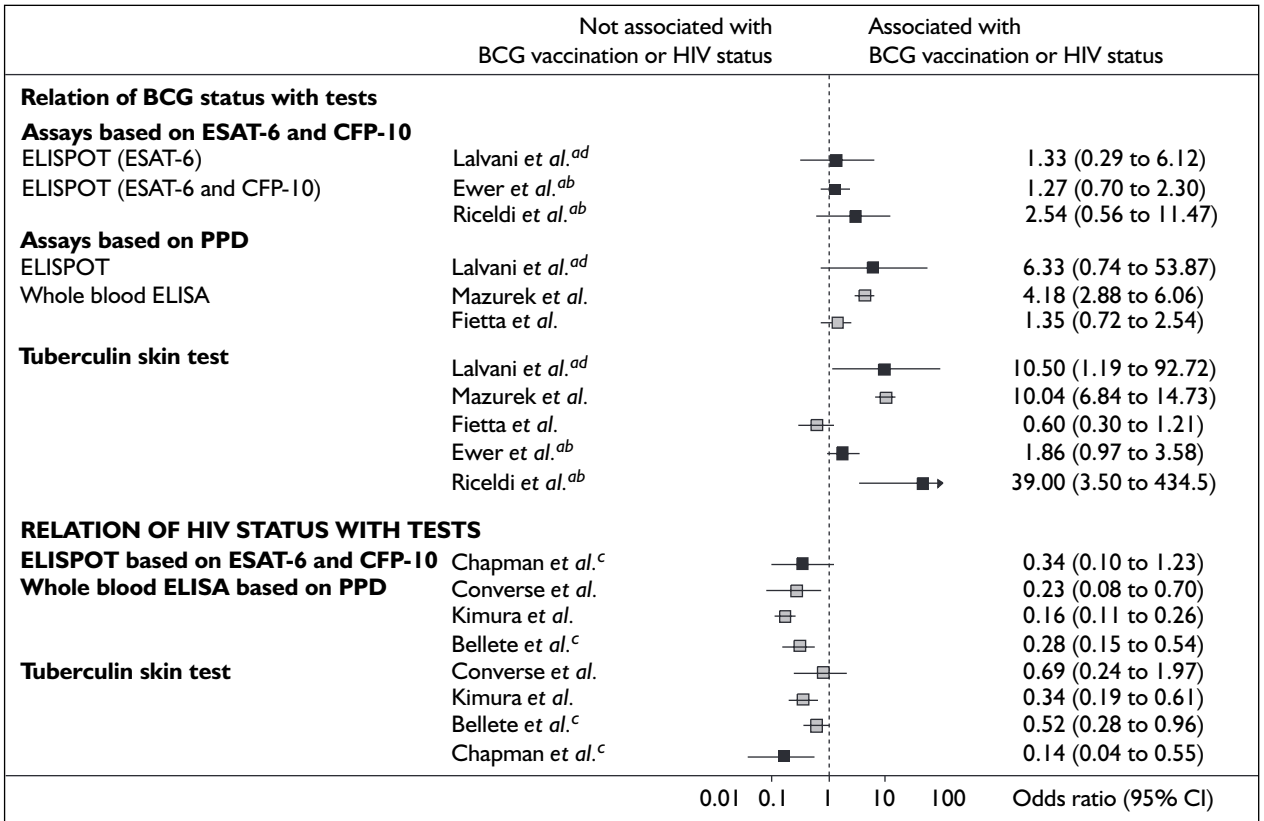
FIGURE 27 *Comparison between interferon-γ assays and TSTs concerning association with exposure to tuberculosis in detecting latent tuberculosis*

and low prevalence countries. RD1-specific antigen-based assays were positively associated with exposure in low TB prevalence countries (ROR 2.07; 95% CI: 95 to 4.53, $p = 0.07$) but negatively associated with exposure in a high prevalence country (ROR 0.66; 95% CI: 0.47 to 0.92, $p = 0.01$). RD1-specific antigen-based assays were more strongly associated with TB exposure than were PPD-based assays (summary ROR 3.65; 95% CI: 1.06 to 12.61, $p = 0.04$). Compared with TST, PPD-based assays were negatively associated with TB exposure overall (ROR 0.5; 95% CI: 0.27 to 0.91, $p = 0.02$), a finding that tended to be consistent in low (ROR 0.73; 95% CI: 0.41 to 1.29, $p = 0.28$) and high prevalence countries (ROR 0.18; 95% CI: 0.02 to 1.67, $p = 0.13$). This meant that RD1-specific antigen-based assays were likely to be more accurate, that is, have false-positive or false-negative results, less often than TST- and PPD-based assays in LTBI in a resource-rich, low TB prevalence setting, an inference supported by four studies of high quality that ascertained exposure adequately.[166,177,178,256] Studies in high TB prevalence countries[258,260] did not show interferon-γ assays (based on RD1-specific antigens or PPD) to be better than TST.

The association of tests with BCG vaccination

To explore which tests can distinguish better between LTBI and BCG vaccination, the association between interferon-γ assays and BCG vaccination was compared with that between TST and BCG vaccination in five studies.[154,158,166,177,178]

Figure 28 shows that PPD-based assays had a significant association with BCG vaccination in 1/3 studies, RD1-specific antigen-based assays had an association in 0/3 studies and TST had an association in 3/5 studies, with 95% CI of OR



CFP, culture filtrate protein; ESAT, early secretory antigen target; PPD, purified protein derivative.
ORs and CIs generated using RevMan 4.2 software.[245]
Solid black squares represent high quality studies; grey squares indicate studies with moderate quality.
[a] Studies with adequate ascertainment of exposure.
[b] Studies investigating an outbreak.
[c] Studies in high TB prevalence country.
[d] RD1-specific antigen-based assay more strongly associated with TB exposure than PPD-based assays in direct comparison.

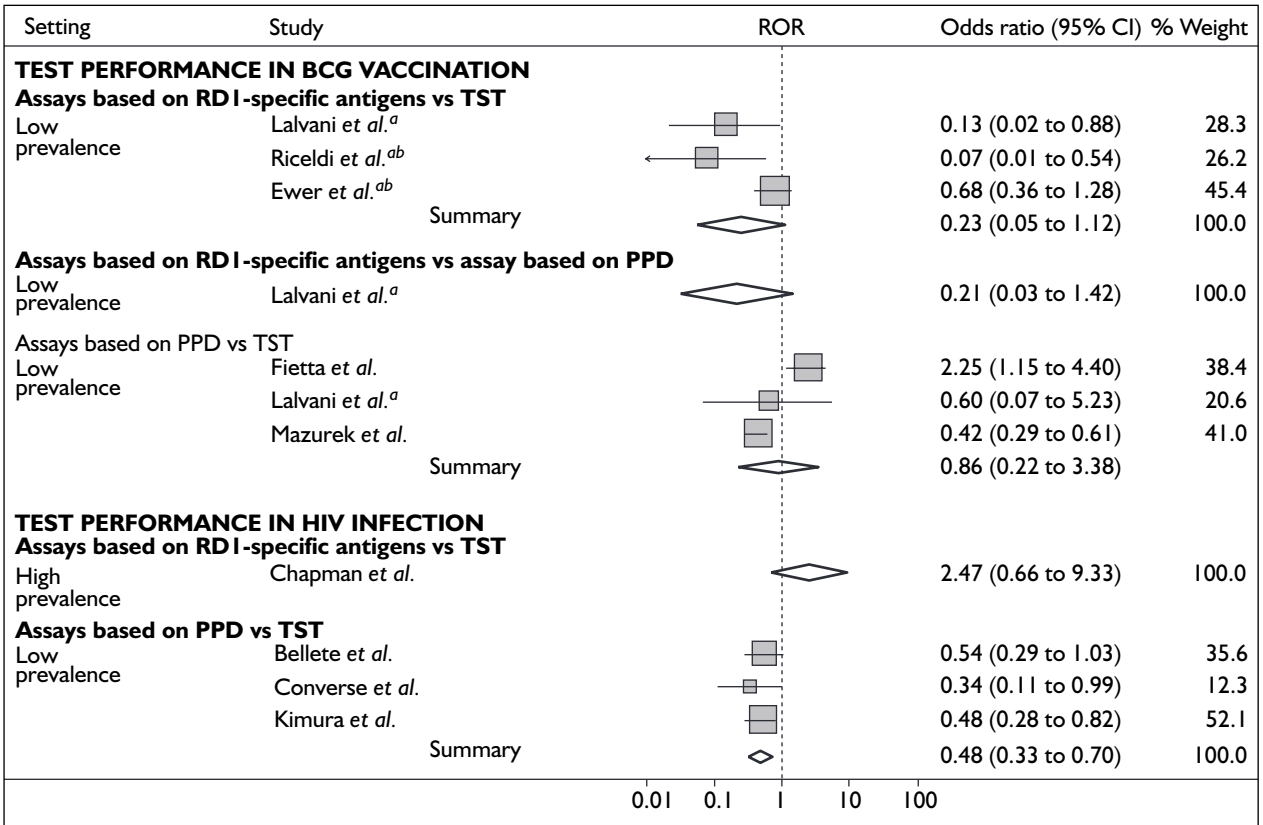
FIGURE 28 *Association of interferon-γ assays and TSTs with BCG vaccination status and HIV status among those exposed to tuberculosis*

excluding 1.0. TST showed some of the strongest associations followed by PPD-based assays.

When the OR point estimates for the tests in *Figure 29* were compared, RD1-specific antigen-based assays were less strongly associated with BCG vaccination than were PPD-based assays (ROR 0.21; 95% CI: 0.03 to 1.42, $p = 0.11$). Compared with TST, RD1-specific antigen-based assays were less strongly associated (ROR 0.23; 95% CI: 0.05 to 1.12, $p = 0.07$), whereas PPD-based assays were as strongly associated (ROR 0.86; 95% CI: 0.22 to 3.38, $p = 0.83$) with BCG vaccination. These findings suggest that assays based on RD1-specific antigens were likely to distinguish better between LTBI and BCG than PPD-based assays or TST, that is, they had false-positive results less often than did PPD-based assays or TST in BCG vaccination, findings supported by studies of high quality in a resource-rich, low TB prevalence setting.[166,177,178]

The association of tests with HIV status

The association between interferon-γ assays and HIV infection was compared with that between TST and HIV infection to assess the impact of anergy due to immunosuppression in four studies.[153,255,257,259] *Figure 28* shows that the OR point estimates of the tests tended to be negatively associated with HIV infection; the negative association was statistically significant in three studies of PPD-based assays and three studies of TST, whereas it was not statistically significant in the one study of the RD1-specific antigen-based assay. *Figure 29* shows that compared with TST, the RD1-specific antigen-based assay was less negatively associated (ROR 2.47; 95% CI: 0.66 to 9.33, $p = 0.18$), whereas PPD-based assays were



PPD, purified protein derivative.
See Appendices 30 and 31 for details for numerical data and *Figure 28* for ORs.
ROR provides comparison between tests: in BCG vaccination, values < 1 indicate less false-positive results; and in HIV infection, values > 1 indicate less false-negative results.
[a] Studies with adequate ascertainment of exposure.
[b] Studies investigating an outbreak.

FIGURE 29 *Comparison between interferon-γ assays and TSTs concerning test performance in relation to BCG vaccination status and HIV infection among those exposed to tuberculosis*

more negatively associated (ROR 0.48; 95% CI: 0.33 to 0.7, $p = 0.005$) with HIV infection. This meant that RD1-specific antigen-based interferon-γ assays were likely to be more accurate, that is, to have false-negative results less often than TST- or PPD-based assays in HIV-positive patients. This finding was supported by one study of high quality that was performed in a resource-poor, high TB prevalence setting.[153]

Discussion

Main findings

This review highlighted that interferon-γ assays have been evaluated in diverse clinical and epidemiological settings in the developing and developed world. The performance of these tests has been compared with TST in groups with high and low ascertained exposures to *M. TB* infection. Assays based on RD1-specific antigens, ESAT-6 or CFP-10, correlate better with intensity of exposure, and therefore are more likely to detect LTBI accurately than TST- and PPD-based assays. An additional advantage is that they are more likely to be independent of BCG vaccination status and HIV status. Studies directly comparing various interferon-γ assays,[166,177,260,261] those with high-quality features[166,177,178,256,258] and those in low-prevalence countries underpinned these findings. Apart from detecting LTBI well in HIV-positive individuals in one study, interferon-γ assays did not perform particularly better than TST in two studies from The Gambia, where only limited data were available.

Strengths and weaknesses of the systematic review

The literature search of this review was exhaustive, but only a relatively small number of eligible studies were identified. Methodological features of the studies included in this review were variable,

but there were sufficient high-quality data to generate inferences.

Among the included studies in this review, assays and tests were conducted and interpreted in different ways; particularly various forms of TST with different doses of tuberculin and different methods of application and interpretation were used. Some of this may be due to the different purposes or populations where different strengths of tuberculin are used and so different cut-offs would be appropriate. A 1.5-mm difference in reaction size may be seen when 10 TU are used compared with 5 TU.[137] Weak doses increase the likelihood of false-negative results and strong doses increase the likelihood of false-positive results. Methods other than the Mantoux technique of intracutaneous injection by needle and syringe deliver inaccurate doses of tuberculin and may cause false results.[29] Only studies in the UK[166,178] used the Heaf test, a method which approximates with the Mantoux test, used in other countries. Although Heaf test reading is usually less precise, the two tests generally correlate well with each other.[70] Different cut-offs were used for positivity of TST, as there is no general consensus on this issue. When interpreting tuberculin reactions, criteria of 5, 10 or 15 mm have been recommended, depending on the clinical situation.[138] Differences in dosages, techniques and interpretation of TST may provide one explanation for the different results observed among the included studies.

In the review, the statistical analysis was limited in a number of ways. First, owing to the small number of studies, the play of chance cannot be confidently ruled out as an explanation for the findings. Tests and exposures to TB were measured in different ways, making it difficult to combine results in a pooled analysis. Differences in exposure status between outbreak investigations and risk comparisons (contacts versus controls) also produce heterogeneity. Analyses used in this review collapsed several exposure categories to generate 2 × 2 tables (see Appendix 27), which inevitably results in loss of information and impinges on the interpretation of summary findings in this review.

A recent non-quantitative systematic review[52] compiled results of studies on interferon-γ assays. It purported to study the sensitivity and specificity of interferon-γ assays, but employed TST results as a gold standard to assess the value of tests in latent TB infection, which is inappropriate. The inadequacies of TST as a reference test mean that such sensitivities and specificities are not true reflections of the performance of the interferon-γ assays. The review also provided results on agreement between interferon-γ assays and TST, which is useful inasmuch as TST has been used for several decades, but again does not give a real indication of test accuracy for the reasons described above. Moreover, the review suffered from a number of methodological deficiencies, including lack of a specific *a priori* research question, limitation of searches to one language, absence of study quality assessment and lack of duplicate checks on data extraction.

Our review has many limitations, but its strength is that it provides the current best summary of the evidence exploring differences between studies' characteristics, quality and results, which leads to a deeper insight into the topic than that afforded by individual studies and non-quantitative reviews.

Interpretation of the review's findings

The ideal test for LTBI would be associated with proven exposure based on known bacillary load and the duration of exposure. However, in practice a compromise surrogate of duration of exposure has been used in studies. The ideal test would also be independent of BCG vaccination and HIV status. How well does interferon-γ comply with this yardstick?

The association of tests with TB exposure

Since airborne spread of *M. TB* is encouraged by close and prolonged contact with a case of infectious TB, the main reason of transmission of infection is the amount of time spent sharing room air with the index case. It is also related to the infectivity of the index case; for example, a patient with heavily sputum smear-positive TB will be more infectious than someone who is smear negative or who is not actively coughing for any given duration of exposure. TST positivity among contacts within a household increases with closeness to the index case; for example, it is higher in individuals sharing a room with the index case than sharing a house.[176] Similarly, the proportion of positive interferon-γ assay results among contacts within a household increases with both closeness to and the length of time spent with the index case.[177,178] The detection of a relationship between exposure and test results depends on the correct classification of individual exposure. If the measure of exposure used in the analysis does not fully capture the burden of infection to which individuals have been exposed, then misclassifications will occur, leading to underestimation of relationships with all test

results, owing to attenuation (or regression dilution) bias. The bias leads to reduced power to detect a difference between the tests. In this review, the interferon-γ assays based on RD1-specific antigens were found to be more closely associated with exposure than those based on PPD and TST. As mentioned in Chapter 3, the measure of association OR is a function of test sensitivity and specificity[246] (one or both of these measures increase as OR increases and vice versa). The approach used in this review cannot directly measure the sensitivity and specificity of the assays for LTBI, but it can enable us to rank the tests according to their expected performance, assuming that intensity of exposure is directly associated with probability of LTBI in the absence of a satisfactory gold standard.

The results from the high-quality studies of outbreak investigations[177,178,256] showed that when there were discordant test results, *M. TB* exposure was more strongly associated with RD1-specific antigen-based assays than with TST. However, there were differences in results amongst the included studies. One possible explanation could be the difference in type of TB organism in the index cases amongst the studies. For example, in one outbreak study[177] the index case had MDR-TB, whereas in the other studies the source case had a fully sensitive organism. The magnitude of improved performance of RD1-specific antigen-based assays also varied from study to study, a finding that may be explained by the differences in timing of testing following exposure as the development of tuberculin sensitivity is time dependent, varying between 2 and 10 weeks after contact.[29] This means that TST, if tested too early, might not have been positive, whereas RD1-specific antigen-based assays might have already been positive. Moreover differences in interferon-γ assay formats and/or prevalence of disease may also explain some of the underlying heterogeneity. The three studies using the T-Spot TB ELISPOT assay and the single RD1 whole blood ELISA study, all conducted in low prevalence countries, had broadly similar results, whereas another ELISPOT assay and the study using an undefined assay, both conducted in a high prevalence country, gave very different results.

High-quality studies comparing PPD-based assays with those based on ESAT-6 and CFP-10[166,177,261] showed that RD1-specific antigens enhanced test performance, an observation similar to another study,[160] that demonstrated improved discrimination between TB infection, most non-tuberculous mycobacteria and reactivity due to BCG vaccination through RD1-specific antigen-based assays. Therefore, the sensitivity and specificity of the RD1-specific antigen-based assays must be higher than that of TST- and the PPD-based assays. In some studies, TST performed better than PPD-based assays. The improved performance of TST in these studies might be explained by the fact that the test was performed under better conditions than usual and was less subject to variation due to the known limitations, for example, operator dependence in routine use.

An interesting finding in this review was the poor performance of interferon-γ assays in a resource-poor country, The Gambia.[258,260] This finding was reproduced in two studies within this country; however, there were no data on relation of exposure to LTBI available from other high-prevalence countries. Since The Gambia has a high TB prevalence, it is possible that a majority of individuals tested in the studies were already latently infected with *M. TB* before contact with the index case, just as >80% of apparently healthy individuals tested in India show positive RD1-specific antigen-based assays.[162] In this situation, positivity of tests might not be consistent with LTBI; it might be a measure of exposure or might be false positive.

Poor test performance in The Gambia could be explained partly by the use of a relatively high cut-off for the ELISPOT assay used here (10 spots compared with five spots used in low-prevalence countries). One Gambian study[260] used a PBMC ELISA with a long incubation period, which might explain the difference in test performance. Another possible explanation for the results observed in a high-prevalence setting is that certain non-tuberculous mycobacteria cross-react with RD1-specific antigen-based assays. For example, exposure to *M. kansasii*, *szulgai*, *flavescens* and *marinum* may contribute to false-positive results in ESAT-6- and CFP-10-based assays.[153] Individuals with intense exposure to non-tuberculous mycobacteria and also *M. kansasii*- or *M. marinum*-infected patients have been shown to have positive responses to ESAT-6 and/or CFP-10.[262] Nevertheless interferon-γ production in these non-tuberculous mycobacterial infections is usually lower than in TB infection.[263] Assays based on ESAT-6 and CFP-10 used in the Gambian studies could have given false-positive results due to these non-tuberculous mycobacteria. Skin test is often positive in patients with exposure to non-tuberculous mycobacteria; however, response to *M. TB* usually gives a stronger reaction.[146] These minor reactions may have been interpreted as

positive either in error due to subjectivity in test interpretation or due to a lower cut-off for TST positivity. Similarly, *M. leprae* ESAT-6 (L-ESAT-6), the homologue of *M. TB* ESAT-6 (T-ESAT-6), was recently identified and recognised by T-cells from individuals who had contact with leprosy and active TB.[264] This might lead to false-positive RDI-specific antigen-based assays results as well as TST in areas of high prevalence of leprosy, although this is a relatively uncommon condition, making this explanation less likely in The Gambia, for example. The two currently available antigens, ESAT-6 and CFP-10, would not be sufficient to discriminate *M. TB* infection accurately from infection with *M. leprae* in regions with a high prevalence of TB and leprosy.[258]

The association of tests with BCG vaccination

This review found that TST, compared with interferon-γ assays, was less able to distinguish between LTBI and previous BCG vaccination. In other studies not eligible for our review, TST- and PPD-based assays, but not RD1-specific antigen-based assays, have been found to be affected by recent BCG vaccination.[160,265] Although the duration of this effect is not known for interferon-γ assays, it was estimated at around 15 years for TST in a recent meta-analysis.[145] A recent study conducted in individuals with no risk factors for *M. TB* exposure in Japan showed that the RD1-specific antigen-based assay was mostly unaffected by BCG vaccination status.[266]

Whereas in developed countries BCG vaccination is not always applied routinely and policy varies between neonatal or school age administration, in developing countries it is generally included in the neonatal/infant immunisation scheme. The degree of BCG-induced delayed-type hypersensitivity varies with country and time.[142] There is considerable loss of hypersensitivity over time in several subtropical and tropical countries as trials conducted in Southern India, Malawi, Sri Lanka and the Solomon Islands have shown.[142,267–269] A recent meta-analysis[145] of the effect of BCG vaccination on TST measurements concluded that TST was of value among BCG-vaccinated individuals in an appropriate clinical setting. For example, if BCG vaccination was more than 15 years previously, it may be ignored as a cause of a current positive TST result, especially if the induration is >15 mm.[145]

The association of tests with HIV status

Screening HIV-infected patients with TST in the presence of cutaneous anergy makes a negative result uninterpretable, leading to underdetection of LTBI.[270] Interferon-γ assays can also lead to false-negative results, especially in patients with advanced HIV infection and low CD4 count, because induced interferon-γ production is much lower in HIV-infected patients with TB infection than in HIV-negative patients,[271] a finding confirmed for PPD-based assays in this review. The review showed that detection may be improved through RD1-specific antigen-based assays. That PPD-based assay did not perform as well in HIV infected people was suggested by three studies, but the better performance of ESAT-6- and CFP-10-based assay was supported by only a single study. It is hard to draw definitive conclusions owing to the small numbers of studies. In a much larger study in active TB,[168] the RD1-specific antigen ELISPOT assay was much less affected by HIV than TST. More research with people with suspected LTBI, particularly in larger numbers of people and stratifying by the CD4 count, is required to examine the trend for performance of different types of assays in HIV-infected people.

Most studies in immunosuppressed individuals have been performed in HIV-infected individuals, so this review only included studies comparing TST and interferon-γ assays against a background of HIV infection. The findings may also be generalisable to other immunosuppressed patients, such as post-transplantation or patients on immunosuppressive medication. Two recent case reports highlight that in immunosuppressed patients RD1-specific interferon-γ assays may be superior to TST.[250,272] In haemodialysis patients, ESAT-6 interferon-γ responses have been shown to be unaffected by uraemia-induced immunosuppression and therefore should be a better marker of LTBI to TST.[273] Ongoing and future studies might clarify this issue.

Implications for practice in resource-poor and resource-rich settings

The assessment of the strengths and weaknesses of interferon-γ assays has implications for their application in various settings and potential impact on TB control. These should be taken into account alongside other characteristics related to practicalities, costs and acceptability of the tests. Use of a blood test rather than a skin test might contribute to TB control. A return visit might not be needed in some settings depending on the test result. The yield of contact investigations should increase. A booster phenomenon will not occur and therefore screening of people who are repeatedly exposed to TB (e.g. healthcare workers) will become possible. Decreased workload in contact clinics will allow overburdened staff to

focus on contact tracing and adherence. It is also possible that the greater flexibility of blood-based tests will increase their overall use in population screening in high-prevalence countries and will increase the number of false-positive cases there, leading to unnecessary chemoprophylaxis and increased toxicity.

Overall, RD1-specific antigen-based assays outperform both TST- and PPD-based assays in all areas of test performance assessed in this review, but not in two studies from The Gambia, a high-prevalence country. The improved performance of the RD1-specific antigen-based assays relative to TST results from improved correlation of assay results with recent TB exposure and lack of cross-reactivity with prior BCG vaccination. In the absence of a gold standard test, these findings strongly suggest that the RD1-specific antigen-based assays are more accurate than TST for diagnosis of LTBI. False-positive results due to prior BCG vaccination avoided with interferon-γ assays can reduce unnecessary chemoprophylaxis and its resulting toxicity. The benefits of interferon-γ assays have not been confirmed in resource-poor settings apart from in HIV infection. These assays might serve as an epidemiological measure for control of TB in view of the above-mentioned advantages in practicality, such as no operator dependence and no need for a return visit. Screening for LTBI with interferon-γ assays will reduce false-negative results in latently infected people with greatest risk of progression in whom TST is often false negative, such as people with HIV infection or iatrogenic immunosuppression. Despite the expense involved, their improved performance may make interferon-γ assays more cost-effective than TST by reducing unnecessary chemoprophylaxis, lessening of number of cases with active TB and decreasing unnecessary use of healthcare resources.

Recommendations for research

The interferon-γ assays should continue to be researched in different epidemiological and clinical settings not only in developed countries, but also in developing countries. Future research in developing countries would give a better sense of whether the lack of superiority of interferon-γ assays shown in our review is an aberrant result or whether these tests are really no better than TST in such settings. Studies should be done not only in countries with high prevalence of TB, but also in those with high prevalence of non-tuberculous mycobacteria and in populations with high BCG coverage.

Studies comparing the interferon-γ assays with TST have been performed in HIV-infected patients; however, there are only limited data available on patients who are immunosuppressed for other reasons. These groups should be targeted for research in addition to HIV-infected groups.

The role of adding more TB-specific antigens to try to improve diagnostic sensitivity further needs to be assessed. Trials to evaluate the performance of the main existing commercial assays (whole blood interferon-γ ELISA and ELISPOT assays) in head-to-head comparisons should be done in both developed and developing countries.

Longitudinal cohort studies to confirm the positive predictive value of interferon-γ assays for subsequent development of active TB should also be performed. Such studies would require careful analysis and interpretation as subsequent development of active disease may be due to infection not contracted at the time of recruitment to the cohort, if the study was done in a high transmission area. These should include sufficiently long follow-up recruiting high-risk groups. They should assess whether changes in strength of interferon-γ responses over time in latently infected individuals can provide an early marker of progression to active TB. Studies that evaluate whether high responses to RD1-specific antigens predict higher risk of developing active disease will be useful.

Chapter 16

Economic aspects to the introduction of new tests for tuberculosis

Introduction

As outlined in the section 'Clinical manifestations of tuberculosis' (p. 2), after exposure to TB, there are three outcomes:

- Infection with tuberculosis [active TB or (ATB)]. This is most often pulmonary, but can be one of the other forms described in the previous chapters.
- Complete clearance by the immune system.
- Incomplete clearance of organisms by the immune system, but with the infection controlled and shut down by the immune system this is latent tuberculosis infection (LTBI). LTBI activates to ATB in among around 10% of cases, with half of these activations occurring within 1 year from infection. The likelihood of LTBI activating to ATB increases if the infected person becomes immune compromised, for example, through the effect of HIV infection, corticosteroid treatment or immunosuppressant drugs.

ATB and LTBI need to be considered separately. In LTBI, the organisms are walled off by the body's immune systems, so any test which relies on the presence of organisms, alive or dead, will not be useful.

Diagnostic tests are not ends in themselves, but only means to better management, including both treatment and prevention. The roles of tests are:

1. Diagnosis – to confirm that the patient has tuberculosis, or to exclude it.
2. To provide antibiotic sensitivities, as a guide to treatment.
3. Identification of organisms. In the context of this review, this refers to whether the organism is the one which is usually the cause of TB in humans (*M. TB*) rather than an atypical mycobacterium such as *M. avium* or *M. chelonei*. Follow-up varies for different organisms. Typing of individual strains of *M. TB* for epidemiological purposes is outwith the scope of this review.

The key economic issues with any new test are first whether it replaces or supplements the older tests, and second the marginal analysis question – if it is better, how much better it is, and whether the extra benefits justify the extra costs, if there are any, since the new test might be cheaper. The cost here is not just of the test, but of the whole clinical pathway, since a more expensive new test might lead to shorter hospital stays. (The usual caveat applies – shorter hospital stays may not release any savings, so there may be no monetary savings, but at least the bed resource could be used for other purposes.) If a new test made the pathway cheaper, the issue is simple (dominant) if it is also better, or one of marginal analysis if it is cheaper but not quite as good.

The benefits of new tests over old ones in the diagnosis and management of TB could include:

- direct benefits to patients through earlier diagnosis of infection, and hence treatment, with less time in ill-health, and possibly reduced mortality
- earlier availability of antibiotic sensitivity, and hence more effective treatment earlier in cases due to resistant organisms, or reduced antibiotic regimens in cases with sensitive organisms
- earlier exclusion of infection in exposed but non-infected people (true negatives with reduction in the duration of prediagnostic anxiety)
- more sensitive diagnosis of latent TB, leading to treatment and hence prevention of some becoming active
- benefits to the health of others, through earlier confirmation of infection in those who might not otherwise be treated; and implementation of control measures including contact tracing; hence reduced spread of disease.

It should be noted that the availability of tests varies amongst countries. This review is concerned mainly with the UK situation, where microscopy, chest X-ray and culture and sensitivity, backed up by reference laboratories, are all available. In some of the world's poorer countries, the only

diagnostic method might be an algorithm based on signs and symptoms, applied by a local health worker; but where prevalence is very high, the health worker may be right much of the time.

The clinical diagnosis pathway for active pulmonary TB

Diagnosis is a cumulative pathway, not relying on a single test, but including:

- the clinical picture, based on history and examination
- chest X-ray
- microscopy, usually of sputum
- other investigations.

In some cases, there will be clinical certainty, for example if there is the combination of clinical signs and symptoms, typical changes on chest X-ray and AFBs in sputum. Confirmation will come from culture. In other cases, there may be a strong clinical impression that TB is unlikely.

The scenarios can be divided into two groups;

- sputum positive for AFBs on microscopy
- sputum negative (or no sputum obtained).

In the first group, if numbers of organisms are large, TB will be diagnosed and treatment will be started. The role of further diagnostic tests is to confirm that the organism is *M. TB*, rather than an atypical mycobacterium.

In the second group, there may still be a presumptive diagnosis of TB, since many patients with ATB are sputum negative. In this review, 21 data sets provided evidence from 3600 patients who were smear negative but had cultures done; 31% of culture-positive specimens were smear negative (range 1–81%). Decisions are made on the whole clinical picture, in effect on informal probability grounds. The key decision in the sputum-negative group is whether to treat on clinical suspicion or not. A decision not to treat would not be irrevocable, since patients could be kept under surveillance and retested. However, there are some groups at higher risk of default from follow-up.

Hence in the microscopy negatives, the aims of further tests will vary according to clinical suspicion:

- high – aim of tests is confirmation. So false-positive results would be a problem because they are likely to be acted on without further investigation
- low – aim of tests is exclusion. So false-negative results would be a problem, because they might end all investigations and follow-up.

In sputum-negative patients, highly sensitive tests such as NAAT may be positive when all others are not. In this review, 25 studies of NAAT accuracy in smear-negative patients (with later confirmed TB) show sensitivity 73.4% and specificity 93.7%. However, the lower the risk of TB, the greater is the likelihood that a positive result is a false positive, even with the good specificity found here.

In the sputum-negative group, the clinician will assess the probability of infection based on:

- Exposure risk.
- Clinical picture mainly from the history (cough, weight loss, fever, night sweats), sometimes aided by findings on examination.
- Chest X-ray, although the classical apical changes may not been seen in the early stages.
- Culture of organisms, which requires far fewer numbers than microscopy. However, culture is positive in only about half of patients thought to have TB in the UK.
- Skin tests, useful mainly in those who have not had BCG. However, about 25% of people with active TB have negative skin tests in the early stages, which increases in the immune suppressed to perhaps 50%. Those who have had BCG immunisation will often show a positive skin test. Induration after BCG is usually under 10 mm but there is no reliable cut-off, and various thresholds have been suggested for different groups.
- The new tests.
- Possibly a therapeutic trial, for 1–2 weeks (or longer – some studies report therapeutic trials of up to 3 months).

The key issue in the sputum-negative group is when clinicians feel confident enough to not treat or to stop treatment. The decision would not be irrevocable as patients could be kept under surveillance and retested if there were clinical signs of disease. However, given a reasonable clinical suspicion of ATB, presumptive treatment is likely. Only if there is a low clinical suspicion in the sputum-negative patients, with the additional tests also showing a negative result, would presumptive therapy not be commenced, raising the possibility of problems with false negatives. In sputum-negative patients, highly sensitive tests such as PCR may be positive when all others are not.

Speed of diagnosis

ATB is relatively slow to progress, the organism only dividing approximately daily. The speed of obtaining confirmation of diagnosis, drug sensitivity and typing of the organism appear likely to have only a limited impact among those with ATB, given the practice of presumptive treatment upon reasonable suspicion while awaiting the results of culture or the new tests. However, there may be some impact upon costs and possibly patient benefits as outlined in the literature review below.

The accuracy of typing of the organism in those with ATB may affect patient impact and overall treatment costs. The importance of accurate typing depends on the prevalence of atypical strains. Although these are relatively high in some overseas countries and consequently in those born within these countries and recently immigrating to the UK, they remain relatively rare in the UK overall (probably less than 5%). Six studies reported a range from less than 1 to 13% of pulmonary TB due to atypical mycobacteria.

Where the speed of diagnosis is likely to have a direct cost impact is within those falsely suspected of having ATB. Among these patients, a more rapid diagnosis ruling out ATB will save money in terms of reducing unnecessary treatment costs. It may also reduce patient anxiety and speed up the correct diagnosis of any other true underlying condition. Given the time lines involved, the differences in the speed of accurate diagnosis appear relatively unlikely to have an impact upon side-effects that may arise in the false positives that are initially treated presumptively for ATB.

An additional complication in assessing the cost-effectiveness of testing suspected ATB cases is that one of the high-risk groups is the homeless. Speed of diagnosis may be important within this group, given the difficulty of follow-up. However, there are also likely to be difficulties in terms of adherence to treatment regimes given their duration. Although a faster accurate diagnosis is likely to increase the proportion of such patients to whom it can be confirmed that they should be being treated before they are lost to follow-up, it is difficult to estimate what impact this will have upon compliance with treatment rates, although it can only increase.

Costs of treatment

The costs of TB treatment are not high; the drug costs are shown in Appendix 32. The 2002 figure for the UK was £1.95 million, of which isoniazid accounted for £230,000 (some of the rifampicin may have been for meningitis due to other organisms such as *N. meningitides* and *haemophilus*). Information available from NHS reference costs related to inpatient treatment for ATB patients indicates that there were 2426 finished consultant episodes for non-elective pulmonary or pleural TB (HRG code D18), at a mean cost of £2219 each. This involved an average hospital stay of 11 days, the 50% range for trusts being £1094–2594. However, these figures may be an overestimate of the true cost as it seems likely that most patients would not be admitted for this length of time: the figures may be skewed by problems in the diagnosis of non-pulmonary disease and by co-morbidities such as HIV.

The clinical diagnosis pathway for latent TB infection

For LTBI, as the immune system has effectively closed down the infection, the current TST relies upon provoking a reaction from the immune system rather than detecting the organism itself. As such, it requires two visits by the patient for a diagnosis, whereas the new tests require only one visit and are not affected by observer error or variations in interpretation. The sensitivity of any test for LTBI is clearly important in cost-effectiveness terms. However, given the possibly low prevalence of LTBI in some of the patient populations under consideration, the specificities of the tests have a major impact upon cost. Specificity is also important in the avoidance of unnecessary patient anxiety and the costs and morbidity associated with treatment side-effects. The analysis of LTBI is complicated by there being no readily available gold standard that will permit the exact estimation of tests' sensitivities and specificities, results between two tests typically being reported as the ROR. Similarly, the estimation of cost-effectiveness depends crucially upon the prevalence of LTBI within the patient group under consideration. The estimation of this is likewise problematic given the lack of a gold standard.

The cost-effectiveness of testing for LTBI will ultimately rest upon the number of ATB cases prevented. Among those correctly treated for LTBI, there is the direct benefit from the number of ATB cases prevented. There is also the public health aspect. Cases of LTBI that develop into ATB and remain undetected will lead to additional infections. The cost-effectiveness of this becomes recursive, some of these newly infected cases clearing the organism from their system, some developing ATB and some developing LTBI.

The detection and treatment of active TB

Review of cost-effectiveness literature

Roos and colleagues[274] measured the cost-effectiveness of introducing PCR to a Nairobi clinic for the detection of ATB. Cost-effectiveness is measured in terms of the cost per correct true positive. The HIV epidemic, with around 10% of adults in Nairobi being HIV positive, is leading to increasing rates of TB smear positive. However, the increase in HIV is also associated with an increase in the rate of extrapulmonary TB cases, and Roos and colleagues report the smear detection rate as being between 30 and 50%. Under current practice, a negative smear may result in a chest X-ray if there are further clinical grounds for suspicion. The reported PCR sensitivity is over 80%, 96% in smear positive and 53% in a smear negative, coupled with a specificity of over 99%, which for the practical purposes of the study is taken to be 100%.

The costs included the cost of testing and the costs of direct treatment of both true and false positives. A significant cost saving is assumed to arise from PCR use, in that it obviates the need for a chest X-ray in smear negative but high clinical suspicion. Overhead costs were assumed to be the same for both test regimes, and so cancel out. Clinic costs were roughly comparable between the two methods. Laboratory costs differed markedly. Smear testing required a significantly higher labour component of US$2.31 per test as against US$0.77 for PCR. Smear testing also involved the use of chest X-ray in smear negative but high clinical suspicion patients, at an average cost of US$1.45 per patient screened. However, these were more than offset by the average cost of laboratory running costs and materials for PCR of US$19.79.

The prevalence of ATB was taken to be 47% within the suspected TB group; 468 out of a cohort of 1000. Current practice of a smear followed by chest X-ray in smear negative but high clinical suspicion detected 347 of the 468 cases. This compares with 374 for PCR, an additional 27 true positives. False positives under current practice numbered 13, as against none for PCR. Including patient travel costs and loss of income, this translated into a cost per true positive of US$40 for current practice as against US$71 for PCR. The implied incremental cost-effectiveness ratio (ICER) for a move from current practice to PCR is consequently US$470 per additional true positive. Excluding patient travel costs and loss of income, this rises to US$557 per additional true positive.

However, as the paper notes, this ignores the costs of false-negative patients who may transmit the disease to others, which would tend to improve the cost-effectiveness ratio given the 8% higher detection rate under PCR. The paper is also only a cost analysis, and so ignores any morbidity associated with the treatment of false positives, and also the morbidity and mortality associated with false negatives and the secondary infections that result from these.

The paper, although interesting, is of limited applicability to the UK setting. Costs will differ markedly. Similarly, the epidemiology of ATB in Nairobi differs markedly from that in the UK. Prevalence rates are likely to be higher within suspected patient groups, and within these patient groups the prevalence of HIV with its associated increase in extrapulmonary TB will also be significantly higher.

Steele and Daniel[275] evaluated a number of different diagnostic strategies for TB in a Bolivian setting with an overall prevalence of 36%:

- smear result alone
- simultaneous smear and chest X-ray
- smear followed by chest X-ray for smear negative
- smear followed by ELISA for smear negative
- ELISA alone.

Costs are applied to these diagnostic strategies, coupled with quality of care values; among those with TB: if treated and cured +1.0, if treated and fail –0.3, not treated –1.0; and, among those without TB: if not treated +1.0, treated with no harm 0.0, treated and harmed –0.1. Unfortunately, the paper reports results against these quality of care values or "expected utilities", although it is honest enough to acknowledge that the relative quality of care values were "assigned arbitrarily on a scale from –1.00 to +1.00". As a consequence, the results are of limited interpretability or generalisability. ELISA appears to fare relatively well within this method.

Lim and colleagues,[276] in a Singaporean study, report that 25–60% of culture-positive patients may return a negative smear at the initial diagnosis. These suspected patients are then typically treated according to clinical suspicion until culture results are attained. From another paper, the authors cite the figure of only 50% of smear negative but subsequently culture-positive patients receiving an initial diagnosis of ATB. The study seeks to identify the relative value of testing

BOX 1 *Base values for the modelling of testing in smear-negative TB*

	Sensitivity (%)	Specificity (%)	Cost (US$)
Clinical assessment	49	90	44
Amplicor NAAT	44	99	59
Bronchoalveolar lavage (BAL)	45	99	287
CT scan	80	60	294
NAAT + BAL	80	95	353
TB treatment	–	–	76

smear-negative patients with the Amplicor NAAT, computed tomography (CT) and bronchoscopy with lavage and/or biopsy (BAL) as compared with the current practice of basing it upon clinical suspicion. In addition to the single test diagnostic strategies, the paper also evaluated the joint test regime of BAL + NAAT and the sequential test regime of NAAT followed by BAL in NAAT-negative patients. There is also a discussion of the use of the more sensitive GenProbe AMTD NAAT. (The GenProbe assay is also modelled with a stated sensitivity of 70%, although no assay cost is reported.) The outcome measure was life expectancy in a cohort of 58-year-old men. Costs included are those of the tests themselves and of treatment for TB.

Base values for the modelling of testing in suspected but smear-negative cases are as shown in *Box 1*.

Immediately striking is the higher sensitivity of clinical assessment as compared with NAAT, although this may in part be due to a lower index of suspicion given its lower specificity. Another critical assumption of the paper is that treatment is not given to the false-negative patients. These patients are assumed not to re-present subsequently with more advanced TB, and as a consequence are assumed to have a mortality risk of 50%. This compares with a mortality risk of 8% in those correctly treated for TB, and it seems high, since it is likely that some would present later.

These assumptions are run for a range of prevalences, with a base value of 5.7% prevalence. Given the assumptions, in particular the reasonably high sensitivity and specificity of clinical judgement, it is unsurprising that the paper concludes that moving away from clinical judgement results in little increase in average life expectancy if the prevalence is 5.7%: less than 1 month's difference between all the test strategies. In particular, Amplicor and clinical assessment showed minimal differences in terms of total cost or life expectancy. Amplicor followed by BAL in Amplicor-negative patients was the only superior strategy to these, but this was more expensive with an ICER (calculated from graph readings) of US$12,600 per additional life-year gained. The other test strategies are dominated. This dominance is reduced and reversed as the prevalence rate is increased, as would be expected given the test's greater sensitivity but poorer specificity. A 50% prevalence sees clinical judgement dominated by Amplicor, it being around US$220 more expensive and resulting in a 2-month lower average life expectancy.

However, the critical assumption of the study, that those deemed negative are discharged, receive no treatment and have a 50% mortality risk, significantly limits the applicability of the study's results. An assumption that patients will not re-present as their TB progresses is unrealistic. It also overstates the importance of the differences in tests' sensitivities, this becoming more serious as the assumed prevalence rate increases. This alone may account for much of the limited difference in the comparison of clinical assessment with Amplicor, given the higher sensitivity assumed for clinical assessment. The sensitivity of Amplicor in smear-negative cases of 44% is within a realistic range, but even only a minor improvement would reverse the results around the comparison of Amplicor with clinical assessment.

Dowdy and colleagues[277] report that incidence rates of TB are beginning to fall relative to diseases from mycobacteria other than TB. The rapid diagnosis from AFB smears does not distinguish between these. Culture readily distinguishes between these, but may take up to 2–3 weeks. The sensitivity of GenProbe AMTD for smear-positive specimens has been reported as between 91.7 and 100% with a specificity of between 99.6 and 100% and the costs of the test are estimated as between US$50 and US$100. Cost savings could result from the early exclusion of smear false positives. Cost savings in the USA might result from the reduction in presumptive TB therapy; recommended

treatment in the USA includes isolation in negative pressure rooms. The public health follow-up is also much less with atypical organisms.

The additional cost of the AMTD assay was based upon US$1200 per kit of 50, coupled with 2.5 hours of technician time at US$25 per hour and some minor additional supplies; this would appear to give a base test cost of around US$100. There are additional costs associated with training of personnel of around US$20 and the holding of positive and negative control samples at around US$2. It is unclear how the paper arrives at a marginal cost per test of US$338, although overall cost-effectiveness is sensitive to throughputs. The daily cost of TB four-drug therapy was US$5.66 and the marginal daily cost of respiratory isolation was a surprisingly low US$28. The base value for the number of smear-positive respiratory specimens that would also be culture positive for TB was 31.4%, and the median delay between GenProbe assay result to culture result was taken to be 6 days.

It has not been possible to reproduce all the results of the paper, given the difficulty in deriving the average test cost. However, the paper reports the base case for routine AMTD testing as costing an additional US$494 (this tallies with a test cost of US$338) per smear positive excluded. The financial savings from each smear false positive correctly identified are US$201. As such, AMTD applied to positive smear results is not cost saving in the base case. However, cost savings are more likely as:

- the prevalence of TB in smear positive patients falls
- the laboratory throughput rises
- the cost per hospital day rises
- the delay between AMTD result and culture result increases.

Although the paper assumes that respiratory isolation is required, which is not the case in the UK, it seems likely that both the cost per hospital day and the delay between assay result and culture result may both be greater within the UK setting. Also, the benefits that arise from avoiding the incorrect treatment of false negatives are not only cost savings: there may be benefits from an early start to correct treatment for the true underlying condition. The benefits from avoiding incorrect treatment for TB in terms of avoiding the side effects from this treatment are likely to be slight, however, given the likely duration of this treatment under either testing regime.

Heymann and colleagues[278] estimate a prevalence of TB in the tested US population of 14%. Their paper is unusual in that it concentrates on the benefits to true positives of identifying the appropriate treatment strategy more quickly, that is, it focuses on sensitivities and sensitivities to drug susceptibility testing. There is no discussion of specificities or of benefits of reducing false positives. The principal benefit appears to arise from rapid radiometric testing having a sensitivity of 98.7% for drug-resistant TB as compared with only 88.5% for conventional drug susceptibility testing. There are also benefits from a more rapid turnaround of tests. The introduction of rapid testing reduces the average time to correct diagnosis from 38.5 to 6.1 days.

Exactly how this is translated into reductions in mortality is unclear, but it appears to rely on a transformation of annual mortality risks, which may tend to overstate the benefits of reducing the duration of inappropriate treatment in drug-resistant TB. The net effect is to reduce the mortality rate from 1.67 to 1.15%. A proportionately similar drop is reported for HIV-positive patients, from 55.3 to 42.1%. The model also indicates that rapid testing will reduce overall health costs per patient by US$272 owing to the higher sensitivity and to the reduction in time to correct diagnosis.

An abstract of Rajalahti and colleagues'[279] paper indicated that the addition of PCR to a smear and culture testing strategy would not be cost saving. However, if the cost per test fell below €97, the test performance was shortened from 4 to 1 day or the proportion of smear-positive patients rose above 4%, the addition of PCR to routine testing would be cost saving. Of more interest, the strategy of applying PCR only to smear positive results was cost saving and dominant. Although this paper cannot be commented on in detail, it underlines many of the conclusions of the literature review above:

- In low-prevalence populations the principal impact of introducing the new assays in terms of costs that are readily measurable arises from the reduction in the numbers of false positives being incorrectly treated for TB.
- The test cost is clearly an important determinant of whether it will be cost saving.
- The daily cost of treatment is also important.
- The time saving from the test as opposed to awaiting the result of culture is critical.
- The prevalence of ATB in the patient group is critical.

- The specificity of the smear test for the patient group is critical.

Discussion and conclusions

A full cost-effectiveness analysis of the benefits of the new tests for ATB would require a combination of the analyses presented above, examining

1. The diagnosis and treatment of false positives:
   (a) the cost savings that might arise from earlier discharge
   (b) the reduction in patient anxiety that may result.
2. The diagnosis and treatment of true positives:
   (a) the sensitivity and specificity of the traditional and new tests to ATB
   (b) the proportion of ATB within suspected ATB patient groups
   (c) the cost, quality of life (QoL) and mortality associated with the correct treatment of typical ATB patients.
3. The typing and treatment of true positives:
   (a) the sensitivity and specificity of the traditional and new tests to atypical and drug-resistant organisms
   (b) the proportion of atypical and drug-resistant patients within the true positives for each test
   (c) the cost, QoL and mortality associated with the correct treatment of atypical and drug-resistant patients, including side-effects
   (d) the cost, QoL and mortality associated with the incorrect treatment of atypical and drug-resistant patients, including side-effects.

The modelling for this would also be required to make major assumptions as to the likely re-presentation rates and the timing of re-presentations among false-negative patients, and from the public health perspective the number of secondary and tertiary infections that would arise from such cases.

In summary, the literature review suggests that the principal benefits in terms of cost-effectiveness may arise from the earlier discharge of false positives and from the higher and earlier correct treatment of atypical and drug-resistant ATB. However, it is not immediately clear that the new tests will have a superior specificity to culture, and the degree to which clinical practice will change is open to debate. The data requirements for a full modelling of this are considerable. Much is currently not available. As a consequence, it has not been possible to undertake a full cost-effectiveness analysis in terms of the cost per quality-adjusted life-year that would arise from the introduction of the new tests. It may be more sensible to attempt to address the first question in terms of the prevalence rate threshold required, below which the introduction of the new tests would save money from the reduction in the duration of treatment of false-positive patients. If this prevalence appears reasonable for certain patient groups, cost-effectiveness appears to follow.

There appears to be limited point in addressing the cost-effectiveness of the new tests in diagnosing and treating those with ATB without addressing the issue of typing. Heymann and colleagues[278] provide a guide to a possible way forward to address the typing and treatment of true positives, but this requires validation within the UK context and a greatly increased data set, including data as to prevalences. There is also the difficulty of quantifying the differential mortality risk that arises from a relatively short reduction in the time to typing of the organism, although Heymann and colleagues also indicate that the new tests may also have greater sensitivity in this regard. However, given the currently relatively low prevalence of atypical and drug-resistant ATB in the UK, the generalisability of studies from the USA that focus on this aspect can be called into question. The lower the prevalence of atypical and drug-resistant ATB, the lower is the cost-effectiveness that arises from being able to identify this rapidly and accurately.

One issue which needs consideration is what would actually happen in practice. Conaty and colleagues[280] reviewed a series of cases to see what changes the results from NAAT tests made in practice. They found that treatment was changed in smear-negative patients not being treated; treatment was started in 17 out of 19. However, the converse did not apply, in that in those smear-negative patients who were being treated, on clinical impression, treatment was continued despite negative NAAT results. However, there were only six of these. Hence the NAAT tests were used as the basis for starting treatment, but not for stopping it. However, the six in whom treatment had been started despite negative smears were drawn from a much larger number of 86 smear-negative patients, so there were presumably compelling reasons for judging them to have active TB.

The detection and treatment of LTBI

In LTBI, the tests which rely on the presence of organisms will not be useful, and we rely on tests

which reflect activity in the immune systems, as described in the section 'Tests for the detection of LTBI' (p. 8). In brief, the current standard test is TST. The new tests rely on interferon-γ assays using encoded proteins (rather than PPD). The newer tests are more expensive in terms of unit cost, but require only one visit, whereas TST requires a follow-up visit to assess the extent of induration in the skin (in reaction to the tuberculin). TST is also prone to variations in interpretation, whereas the new tests are not.

Review of cost-effectiveness literature

The cost-effectiveness literature on testing for LTB unearthed only one paper that went beyond a simple description of the issues. Rose[281] provides a Markov model of screening for LTBI with TST within various different patient groups: 3-year-old children exposed to high-risk adults and in a number of groups of 3-year-old adults. Prevalence rates of LTBI within these groups for modelling are as follows:

- children exposed to high-risk adults: Not clearly stated
- household contacts with active cases: 25–50%
- recent immigrants from high-prevalence countries: 10–40%
- residents and employees of high-prevalence prisons: 10–40%
- intravenous drug abusers: 10–20%
- high-risk medical conditions: 2–4%

The prevalence among those with high-risk medical conditions is taken to be that of the general population, for whom screening has been discontinued. Cost-effectiveness in these patient groups may arise owing not to their prevalence of LTBI, but to their increased risk of LTBI developing into ATB. Their relative risk of this compared with the general population with LTBI is as follows:

- HIV: 9.7–170.3
- end-stage renal disease, transplantation: 19.0
- end-stage renal disease, dialysis: 1.6–16.0
- gastrectomy: 5.0–6.8
- recent major weight loss: 1.8–2.1
- silicosis: 1.5–32.8
- diabetes: 2.0–4.1
- leukaemia or lymphoma: 1.0–35.0
- intravenous drug users: 3.2–19.2

The structure of the model is not clearly given, but it is stated as using a 10-year public health perspective including secondary and tertiary infections. Transmission rates by those with ATB are uncertain. US data since 1950 suggest 0.5–2.2 people being infected per active case, although modelling based on US national data suggests up to 3.5 infections per active case. Old Dutch data from 1921 to 1938 suggest an infection rate of 13 from each active case, although the applicability of these data to the current setting is unclear. For secondary and tertiary infections, half are identified and treated with preventive therapy, presumably before developing advanced ATB. Results are reported in terms of the numbers that need to be screened to:

- detect one case of LTBI
- prevent one case of ATB
- prevent one death from ATB
- increase life expectancy.

However, the population risk of TB was assumed to be the prevalence of infection as measured by the TST multiplied by the infection rate among those known to be infected. These rates are taken from two different papers, and it is unclear whether this implicitly assumes a TST sensitivity of 100%. This suspicion is heightened by the reported number that need to be screened to detect one case of LTBI. For instance, in the household contacts with active cases group, the number needed to screen to detect one case of LTBI is stated as being between two and four, which corresponds to a 100% sensitivity and a prevalence of 25–50%.

Isoniazid preventive treatment is reported as reducing the risk of developing ATB by 65% for a 6-month treatment and by 75% for a 12-month treatment. This is similar to the rate reported in Jasmer and colleagues[282] of 60. They also note a wide range of 25–92%, with 92% for good adherence.

The aspect of the paper[282] that raises most concern is that it does not discuss the sensitivity or specificity of the TST. There is no discussion of the number of false positives and how these may differ between the different risk groups, or of the impact of incorrect treatment in false positives. All those screened with TST and having a positive test result appear to be assumed to be at an increased risk of developing ATB. This may be reasonable and reflect the joint probability of having LTBI and developing ATB among those with a positive TST, but it appears possible that the modelling has assumed a 100% specificity for the TST. This seems all the more likely given that the data for the modelling are drawn from disparate sources, rather than a single, cohesive study.

BOX 2 *Advantages and disadvantages of detecting LTBI*

Option	Advantages	Disadvantages
Treat all with LTBI	Active disease prevented Possibly, prevention of spread	Cost (but drugs cheap) Side-effects of drugs
Surveillance and treat only when active	Reduced cost Fewer side-effects	Illness from ATB Possible spread to new cases Danger of loss to follow-up

All the groups analysed benefited from screening, and those identified as LTBI benefited from treatment in terms of increased life expectancy. Quantifying the benefits is difficult, however, given the wide range of values within the literature for the various risk factors. The benefits of screening are similarly reduced if the group in question is unlikely to complete treatment. However, the benefits from treatment in terms of life expectancy are small on average.

Average life expectancy increases in all groups, in that the benefits of treatment do appear to outweigh the possible side-effects in terms of mortality. However, the average gains are extremely small, and within the groups being screened the maximum average gain per patient screened is typically a fraction of a month. Only among those who are HIV positive and among intravenous drug users do these average maximum gains rise above 1 month. Also, it must be stressed that these are the maximum possible gains, which is probably more a reflection of the data uncertainties; the minima remain as fractions of a month gain in life expectancy. Indeed, the minima are typically less than an increase of 1 day in average life expectancy, only HIV-positive patients rising above this with a minimum of 5.4 days' increase in life expectancy.

The paper does not address the morbidity side-effects of treatment. This is problematic given that there will be a number of false-positive patients who undergo unnecessary treatment and experience these side-effects. As the paper does not seem to address the problem of false positives, the extent of this is unclear.

Discussion and conclusions

Screening for LTBI implies that a positive diagnosis will result in treatment for LTBI. However, there are at least two options, with advantages and disadvantages (*Box 2*).

The cost-effectiveness of screening for LTBI within defined patient groups will consequently depend upon a number of factors:

- the proportion with LTBI within the patient group
- the cost, sensitivity and specificity of the test regime for LTBI within the patient group
- the likelihood of those within the patient group with untreated LTBI developing ATB
- the cost of treatment for LTBI and its effectiveness in reducing the risk of ATB within the patient group
- the side-effects of treatment for LTBI and their cost of treatment
- the transmission rate by those within the patient group who develop ATB
- the cost and effectiveness of the detection and treatment of ATB.

Regarding the first five points, among the general population who are not high risk, the low proportion with LTBI in the UK coupled with a non-elevated risk of developing ATB means that it is generally agreed that screening for LTBI is undesirable. The costs, inconvenience and side-effects from the treatment of the large number of false positives that would result are not balanced by the gains arising from the treatment of those correctly diagnosed as having LTBI.

A high test sensitivity is clearly desirable, but this has to be set against the cost of the test. From first principles, it can be argued that the greater the likelihood of LTBI developing into ATB among a patient group and the greater the reduction in this risk from the treatment of LTBI, the greater should be the willingness to pay for a higher test sensitivity. The detection of true positives is more likely to be cost-effective, implying that tests with a higher sensitivity are more likely to be acceptable even if they cost more. The importance of test sensitivity for cost-effectiveness will also be linked to the prevalence of LTBI in the patient group. In contrast, the importance of test specificity increases as the proportion of the patient group with LTBI declines and as the cost and side-effects from incorrectly treating those without LTBI increase.

The appropriate index of suspicion for a test regime is correspondingly complex, and likely to

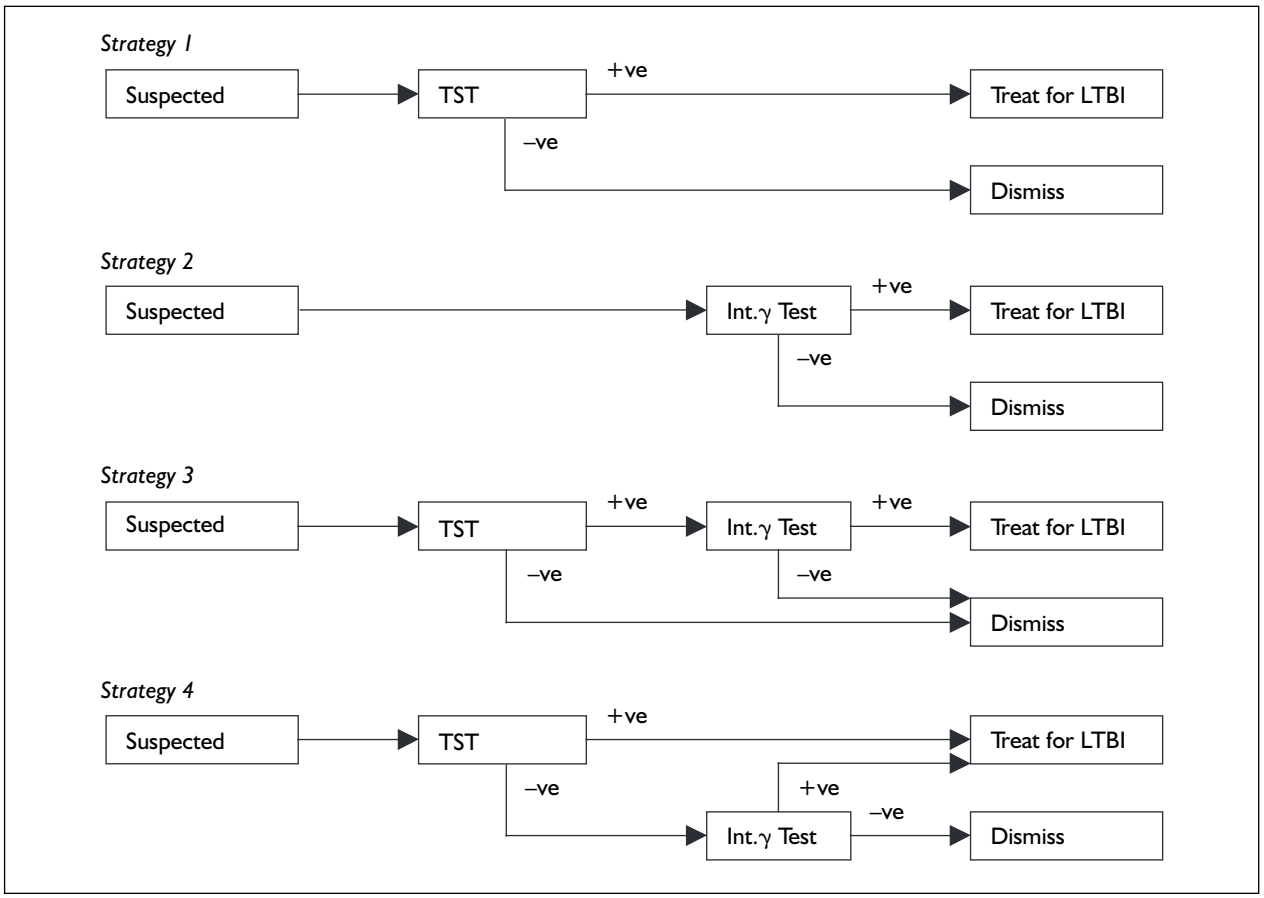


FIGURE 30 *Alternative strategies for screening for LTBI*

differ between different patient groups. This index of suspicion can be altered for the individual tests under consideration, for example, by altering the cut-off for the size of skin reaction under TST that is labelled a positive. Sensitivities and specificities could also conceivably be altered through test combinations, as outlined in *Figure 30*. The choice of strategy would depend on whether those being tested had had BCG (or BCG in the last 15 years), and also on the optimum balance between sensitivity and specificity, which would vary with expected prevalence.

An aspect of the new tests that is not addressed in *Figure 30* is whether they can type the organism and provide information as to drug sensitivity within LTBI. This would increase the cost-effectiveness of all the strategies involving interferon-γ testing, and would tend to promote strategy 2 or strategy 3 above the others.

The desirability and feasibility of multiple test regimes may be questionable. Applying tests sequentially as in *Figure 30* implies a further delay to treatment under the third strategy. This may not be problematic in terms of disease prognosis, given the limited and slow progression from LTBI to ATB, but it may be problematic in terms of loss to follow-up among some patient groups. The shorter delay from testing with interferon-γ as opposed to TST is already being cited as an advantage of the technology for some patient groups, such as the homeless. A joint, concurrent testing regime could be envisaged, although this would increase the cost of testing somewhat over the single test regimes, and over the sequential joint regime where Interferon-γ acts as a further gatekeeper to treatment for LTBI.

However, joint testing regimes could be used to alter the emphasis on sensitivity as against specificity, tending to emphasise specificity as in strategy 3 or emphasise sensitivity as in strategy 4. The degree to which this could be achieved through altering the index of suspicion of individual tests is unclear, although this could be of lower cost and more convenience than the joint testing outlined above.

The performance of the individual tests for LTB in terms of their sensitivity and specificity can be broadly defined according to three patient groups:

- immune competent without BCG
- immune competent with BCG
- immune suppressed.

In the immune competent group without BCG, the TST gives high sensitivity to LTBI (ATS guidelines 1997[102]). Specificity is high: 99% in populations which have had no other mycobacterial exposure or BCG. However, if prevalence of TB is very low, the positive predictive value (PPV) will be low. For example, if prevalence of TB infection is 1%, the PPV is 0.16.[102] Most positives will be false positives.

One problem with strategy 1 is that it assumes that immune competence is known, which may not always be the case.

Previous BCG (usually regarded as being within the 15 years) can result in false-positive reactions to TST in those who are LTBI negative. The positive proportion depends on definition, in terms of breadth of induration. The effect on test performance appears to fall through time. Jasmer and colleagues[282] report that 8% of those with BCG at birth give a positive response to TST 15 years later. The paper does not report whether this 8% is both true and false positives or only false positives, or the background prevalence of LTBI in the population tested. Mazurek and colleagues[158] found the likelihood of a positive TST result but negative interferon-γ to be seven times greater among those with BCG than those without (all were over 18 years old). The false-positive rate arising from BCG among interferon-γ assays is lower than that for TST, although it should be noted that some of these assays use PPD; those based upon encoded antigens (the secreted antigen is not present in BCG) distinguish between BCG and LTBI somewhat better than PPD-based assays (Master's thesis,[254] although based upon OR and ROR rather than sensitivity and specificity dominance *per se*). As a consequence, the encoded protein interferon-γ assays are more likely to be cost-effective in those having had BCG than those who have not owing to their relatively better sensitivity.

In contrast, immune suppressed patients appear less likely to react to TST, so worsening the test sensitivity. This can in part be compensated for by reducing the index of suspicion and the size of reaction required for a positive result, as reported in Jasmer and colleagues,[282] although there will be a corresponding specificity penalty. The sensitivity of both interferon-γ assays and the TST may be reduced in immune suppressed patients (Frieden and colleagues[7] reporting for HIV; it is not explicit whether this applies to all interferon-γ assays). However, although both tests' sensitivities may be compromised, it again appears that interferon-γ assays may perform better than TST under such circumstances, with assays based on encoded antigens again appearing more accurate than PPD-based assays (Chapter 15). As in the above considerations of BCG, the interferon-γ assays are more likely to be cost-effective in those who are immune suppressed owing to their relatively better sensitivity. This likelihood is further increased by the greater likelihood of LTBI developing into ATB.

A cost minimisation approach in those who had had BCG, and were deemed to be immune competent, would be to use TST as the initial test, as in strategy 3 (*Figure 30*), and to use the interferon-γ test only in those who were TST positive, to distinguish false positives from assumed true positives.

Strategy 4 (*Figure 30*) would be for those who had not had BCG, and were deemed immune competent. Here, the greater sensitivity of the interferon-γ tests is used in those who are TST negative, but they also avoid the false-positive TST reactions among those who were responding as the result of exposure to mycobacteria other than *M. TB*.

A relatively simple model of the effectiveness of these testing strategies could be developed, following the lines of Rose[281] in allowing for less than perfect specificities but also allowing for less than perfect sensitivities (*Figure 31*).

Given the discussion of the differing accuracy of individual tests within different patient groups, the numbers going down the initial different arms of true positive, false negative, true negative and false positive will differ markedly between groups and test strategies. These numbers will also be determined in large part by the prevalence of LTBI within each patient group. The other critical difference between the patient groups will be in terms of the likelihood of developing ATB from LTBI. Certain groups may have notably different likelihoods of being ATB and remaining undetected as opposed to being picked up as a suspected ATB case. To this must be added a probably higher likelihood of providing secondary new infections within each period. Death rates from ATB will be higher among the immune suppressed.

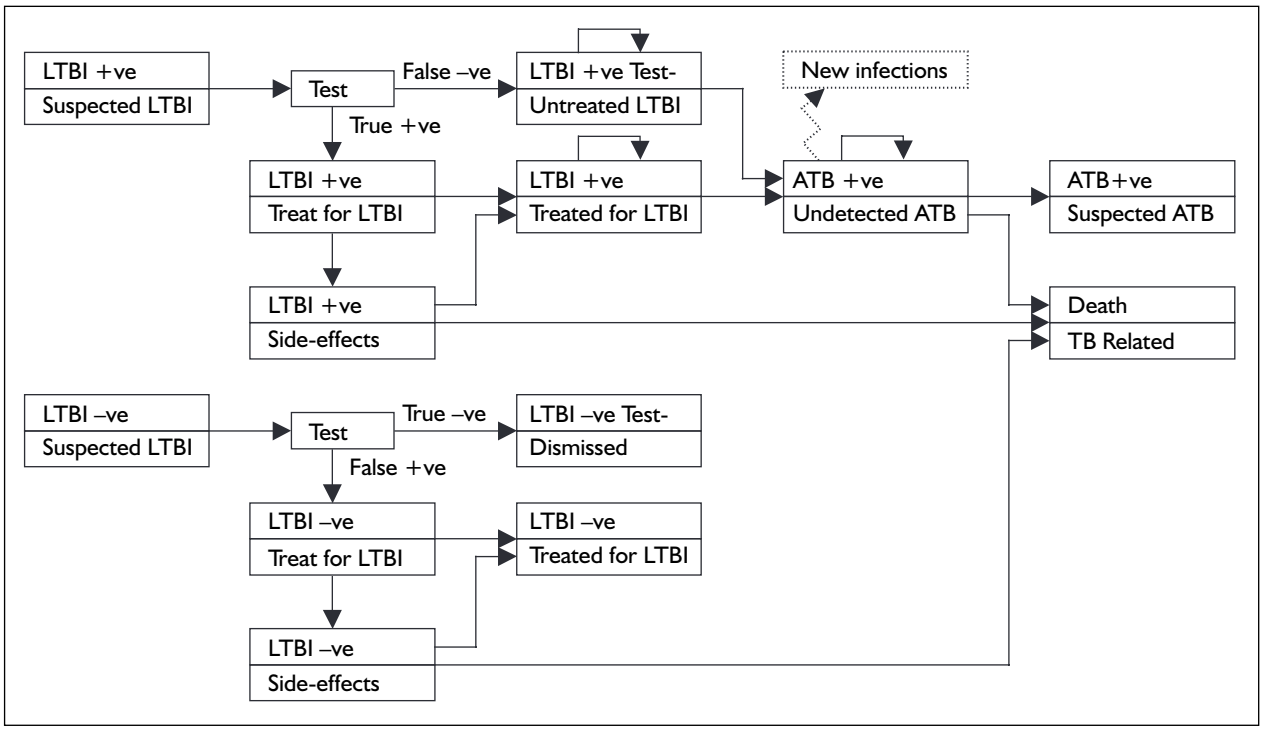


FIGURE 31 *A simple model of the effectiveness of alternative screening strategies for LTBI*

The death rate from side-effects from treatment is extremely low within the general population (reportedly of the order of 0.02 per 1000 treated[281]), to the extent that they could be ignored with limited effect upon overall results. However, it is not clear whether this also applies to all immune suppressed groups. The side-effects that may occur and the treatment of these may also differ.

Patients exit the model in *Figure 31* either through death, inappropriate treatment for LTBI or developing ATB and being suspected as such within the health system. The first two are appropriate end-points for the model. The last relates to the previous section and suggests the need for a full model of both LTBI testing and ATB progression including testing. However, in the light of the discussion as to the feasibility of full modelling of ATB testing and progression, this is unlikely to be feasible in the immediate future. The entire rationale for testing for and treating LTBI is to avoid ATB. If the cost-effectiveness of testing for ATB among suspected ATB cases in terms of cost per quality-adjusted life-year (QALY) is currently difficult to estimate, this is doubly so for testing for LTBI. In order to run the model, it may be possible to assign optimistic and pessimistic values to the patients that develop ATB in terms of costs and QALYs, although how these might be arrived at is not immediately apparent. As a consequence, any cost-effectiveness analysis for testing for LTBI in terms of costs and QALYs is likely to be partial, and potentially misleading.

It may be more appropriate at present to limit modelling and the data collection required for it to a more simple cost–consequence analysis. The simplest of these would be to itemise the test costs incurred under the different strategies and report this against true positives, false negatives, true negatives and false positives that would occur within the different patient groups. A single outcome measure of correct diagnoses could be used, although this would equate the value of true positives and true negatives. The principal uncertainties that would need to be addressed for this modelling revolve around the lack of a gold standard:

- the absolute sensitivities and specificities of the tests, as opposed to their relative performance as measured by the ROR
- the proportions with LTBI within each patient group.

Neither of these are readily resolvable, and modelling would probably have to be run over a range of plausible values.

However, extending the modelling slightly to encompass the treatment for LTBI among both

true and false positives coupled with the possible side-effects from treatment should be relatively simple. Indicative data as to the types and costs of side-effects should also be available, although there may be some problems estimating this for each of the immune suppressed groups. This extension to the modelling would encompass the potential cost savings from an improvement in specificity that may arise from the new tests, in addition to highlighting the reduction in harm that may occur from fewer false positives being treated incorrectly for LTBI. A cost–consequence analysis could be presented, highlighting any changes to the rates of side-effects in false-positive patients but concentrating upon changes to costs and the single outcome of correctly treated true-positive patients. An indication of the likely number of ATB cases that would develop and the secondary infections that might result could also be based on the effectiveness of treatment for LTBI values, LTBI progression values and secondary infection values reported in the literature, such as by Rose,[281] although the applicability of these to the UK setting may require verification.

Chapter 17
Discussion

In the discussion which follows, it should always be remembered that diagnosis will rarely be made on the basis of a single test, but rather on the complete clinical picture, starting with history and physical examination; going on to non-laboratory investigations such as chest X-ray, taking probabilities into account (such as level of exposure), and then a sequence of laboratory tests starting with microscopy.

Summary of key findings

NAAT tests

Overall, we found NAAT test accuracy to be far superior when applied to respiratory samples as opposed to other specimens. Although the results were not statistically significant, the AMTD test appears to perform better than other currently available commercial tests. The better quality in-house studies were, for pulmonary TB, much better at ruling out TB than the commercial tests (higher sensitivity), but were less good at ruling it in (lower specificity). Given that these tests are in no way standardised and cover a wide range of different target genes and procedures, it is not possible to recommend any one over another owing to a lack of direct test comparisons.

The specificity of NAAT tests was still found to be high when applied to body fluids, for example for TB meningitis and pleural TB, but sensitivity was almost uniformly poor, indicating that these tests cannot be used reliably to rule out TB but are of value in 'ruling in' a diagnosis with compatible clinical and/or radiological findings. The main problem appears to be with the sensitivity of the tests, presumably due to difficulties in obtaining fluid samples with sufficient mycobacteria to allow amplification. High specificity estimates suggest that NAAT tests should in fact be the first-line test for ruling in TB meningitis, but that they need to be combined with the results of other tests in order to rule out disease.

Evidence for NAAT tests for the detection of other forms of TB is significantly less prolific than for those above. Although there is some evidence for their potential use in lymphatic TB and genito-urinary TB, there was insufficient data from the studies examined to draw any conclusions or make any recommendations.

Phage tests

Although the body of evidence is small, the phage tests have been shown to have high specificity in pulmonary TB but it is not clear whether these tests have sufficiently high sensitivity in smear-negative samples to recommend their routine use in practice. As discussed previously, they do have the potential advantage over NAAT tests that they detect only viable mycobacteria. Phage and some forms of NAAT are able to yield antibiotic sensitivities within 2–3 days, although rapid antimicrobial resistance was not systematically examined in this review.

Adenosine deaminase tests

There is no evidence to support the use of ADA tests for the diagnosis of pulmonary TB. However, there is considerable evidence to support their use in pleural fluid samples for diagnosis of pleural TB, where sensitivity was very high, and to a slightly lesser extent for TB meningitis. In both pleural TB and TB meningitis, ADA tests had higher sensitivity than any other tests. Further research is required to determine the specificity of this test with regard to non-tuberculous pleural effusions of infectious or inflammatory aetiology. Of all the tests evaluated for peritoneal or pericardial TB, ADA appeared the most promising; however, few studies of other tests were identified and further research is needed.

Interferon-γ

We found the assessment of interferon-γ levels in pleural fluid samples to be very promising for the detection of pleural TB. Diagnostic sensitivity was high, but further studies, including patients with infectious and inflammatory non-tuberculous pleural effusions, are required to determine the likely specificity of these tests in routine practice for evaluating patients with suspected pleural TB.

Anti-TB antibody tests

Serum anti-TB antibody test performance was universally poor, regardless of type of TB, although some studies did show high specificity.

Fully automated liquid culture

We found fully automated liquid culture methods to be superior to culture on solid media in terms of their speed and their precision for the detection of isolates compared with solid media. Although the BACTEC 460 radiometric method also has the same benefits, it is radiometric and therefore requires disposal of radioactive waste and also requires more staff time compared with the fully automated methods. The fully automated methods also had higher contamination rates than the BACTEC 460, but data on contamination rates may be somewhat outdated, as recent developments including the addition of antibiotics to the specimen vials appear to have reduced contamination. There were insufficient data to determine if antibiotic addition increased the time-to-culture for specimens.

Interferon-γ assays for detection of latent TB infection

Interferon-γ blood tests based on RD1-specific antigens, ESAT-6 or CFP-10, correlate better with intensity of TB exposure, and therefore are more likely to detect LTBI accurately, than TST- and PPD-based assays. An additional advantage is that they are more likely to be independent of BCG vaccination status (owing to higher specificity as the key proteins are not produced by BCG) and HIV status (owing to higher sensitivity). Interferon-γ assays did not perform particularly better than TST in two studies from The Gambia, where only limited data were available.

Explanations for variations amongst study results

Tests for active TB infection

The large number of available studies for NAAT tests in pulmonary TB meant that we were best able to detect reasons for variation in study results amongst this group of tests. Overall, we found the main explanatory factors for the variability to be the reference standard used, whether the study was laboratory or hospital based, and the use of blinded test interpretation, suggesting that study design-related factors appear to have more impact on study results than patient- or setting-related factors. Blinded interpretation of both index and reference tests was associated with lower accuracy, providing further weight to calls for improved study design and reporting.

The overall reporting of studies was poor, and was similar across all test types, making the true accuracy of the tests difficult to estimate with certainty, especially when small numbers of studies are available. Many of the studies focused on technical comparisons with existing methodologies, which was very useful, but few studies used our designated 'ideal' reference standard, that is, culture plus high clinical suspicion with or without other diagnostic interventions, thereby making difficult an assessment of how the tests truly perform in clinical practice for pulmonary specimens.

Tests for latent TB infection

The results from the high-quality studies of outbreak investigations showed that when there were discordant test results, *M. TB* exposure was more strongly associated with RD1-specific antigen-based assays than with TST. However, there were differences in results amongst the included studies, which may potentially be explained by differences in interferon-γ assay formats and/or prevalence of disease. This review found that TST, compared with interferon-γ assays, was less able to distinguish between LTBI and previous BCG vaccination. The review showed that detection of TB infection may be improved in HIV-infected people through RD1-specific antigen-based assays.

Strength and limitations of the review

The systematic nature of the review means that we are likely to have identified the majority of the published studies. The literature search was comprehensive, using a wide range of electronic databases and relatively broad search terms, such that all of the indexed literature should have been picked up. Two reviewers were involved at every stage in the review procedure, such that mistakes due to human error should be limited. A quality assessment tool that has been developed according to scale development principles was adapted and applied to each of the included studies.

Empirical evidence suggests that studies with significant or favourable results are more likely to be published than those with non-significant or unfavourable results.[283] There is as yet no evidence for the degree of publication bias likely in the field of diagnostic tests, but there is no reason to believe that it will be any better than in studies of therapeutic interventions. It is therefore possible that we have missed a proportion of English-language studies. Time and resource constraints meant that we could not assess any foreign language papers for inclusion.

Nevertheless, we did identify a large number of studies, particularly for pulmonary TB and pleural TB. By restricting the studies included only to those using a case series or cohort design, we have provided a better indication of how well the tests will perform in practice, given the potential biases associated with case–control studies.[183]

Chapter 18

Conclusions and recommendations

Implications for practice

ATB is diagnosed by a combination of clinical judgement based on a review of symptoms and clinical signs, radiology and laboratory testing. In practice, for pulmonary TB, clinical judgement often has higher sensitivity (but not necessarily higher specificity) than laboratory diagnosis of TB. Culture-based identification has been the accepted 'gold standard' for the laboratory diagnosis of TB, and indeed some national reporting systems only permit the recording of bacteriologically confirmed cases for the reasons given above. In practice, a suspicious clinical history and chest X-ray are often used as the 'clinical proof' of pulmonary TB regardless of laboratory analysis, so that TB may be over-diagnosed (picking up old non-active TB lesions) and also under-diagnosed (missing true cases). The NAAT tests provide a reliable way of increasing the specificity of diagnosis (ruling in disease), but sensitivity is too poor to rule out disease, especially in smear-negative disease where clinical diagnosis is equivocal.

Diagnosis treatment decisions for pulmonary TB, for example, need to take three issues into account (*Figure 32*):

1. The prevalence of disease in the different levels clinical suspicion (which will represent a spectrum; the division into three is just for convenience).
2. The proportion of patients which we would feel comfortable treating erroneously but unknowingly. Where clinical suspicion is high, it is likely that over half of the patients will have ATB, but in the low clinical suspicion group, most will not have TB; the percentages who would be treated incorrectly if we relied on, for example, NAAT positivity would be very different.
3. The 'full picture effect' – decisions would be made on a sequence of investigations, so a positive NAAT test would always be subject to interpretation in the light of prior evidence.

Overall, NAAT test diagnostic sensitivity is still insufficient to provide a reliable, rapid rule-out test for paucibacillary pulmonary TB, especially for smear-negative disease, where the clinical need is greatest owing to the poor sensitivity of conventional diagnostic tests. The NICE guidelines on TB,[9] recommend that molecular methods be used only where rapid confirmation of TB diagnosis in a sputum smear-positive individual would alter the patient's care (presumably referring to cases where MDR-TB is suspected), or before conducting a large contact-tracing initiative.

Our findings indicate that for pulmonary, smear-positive TB, NAAT tests should be used regardless of degree of clinical suspicion, to distinguish between *M. TB* and NTM or to identify the NTM in question, or to identify MDR-TB. Distinguishing *M. TB* from atypical mycobacteria is important, since the public health follow-up is different (full follow-up with contact tracing is needed if *M. TB* is found, but not if, for example, *M. avium*). Public health follow-up can be time consuming and difficult, for example, if exposure has been on a long-haul air flight. Although not covered in our review, NAAT tests are used for drug susceptibility testing if there is suspicion of drug resistance, perhaps based on country of origin or previous incomplete treatment. Antibiotic resistance, however, has also been reported in UK-born children with no previous treatment.[284] Levels of multiple drug resistance are low in the UK, but there have been outbreaks due to resistant organisms in London, where MDR-TB is more common.[285]

It is in smear-negative disease that NAAT tests could have the greatest impact, particularly in terms of ruling in disease, as the lower sensitivity makes ruling out disease more difficult. For smear-negative disease, appropriate use of a NAAT test will depend on the associated degree of clinical suspicion. *Table 80* presents some calculations of the effect of a positive or negative NAAT test result (using the sensitivity and specificity estimates established in *Table 7*) on the probability of TB being present in smear-negative patients according to different levels of clinical suspicion. These estimates show that if we assume an 'acceptable' level of unnecessary treatment of false-positive cases of 10%, the index of clinical suspicion needs to be at least 40% before a

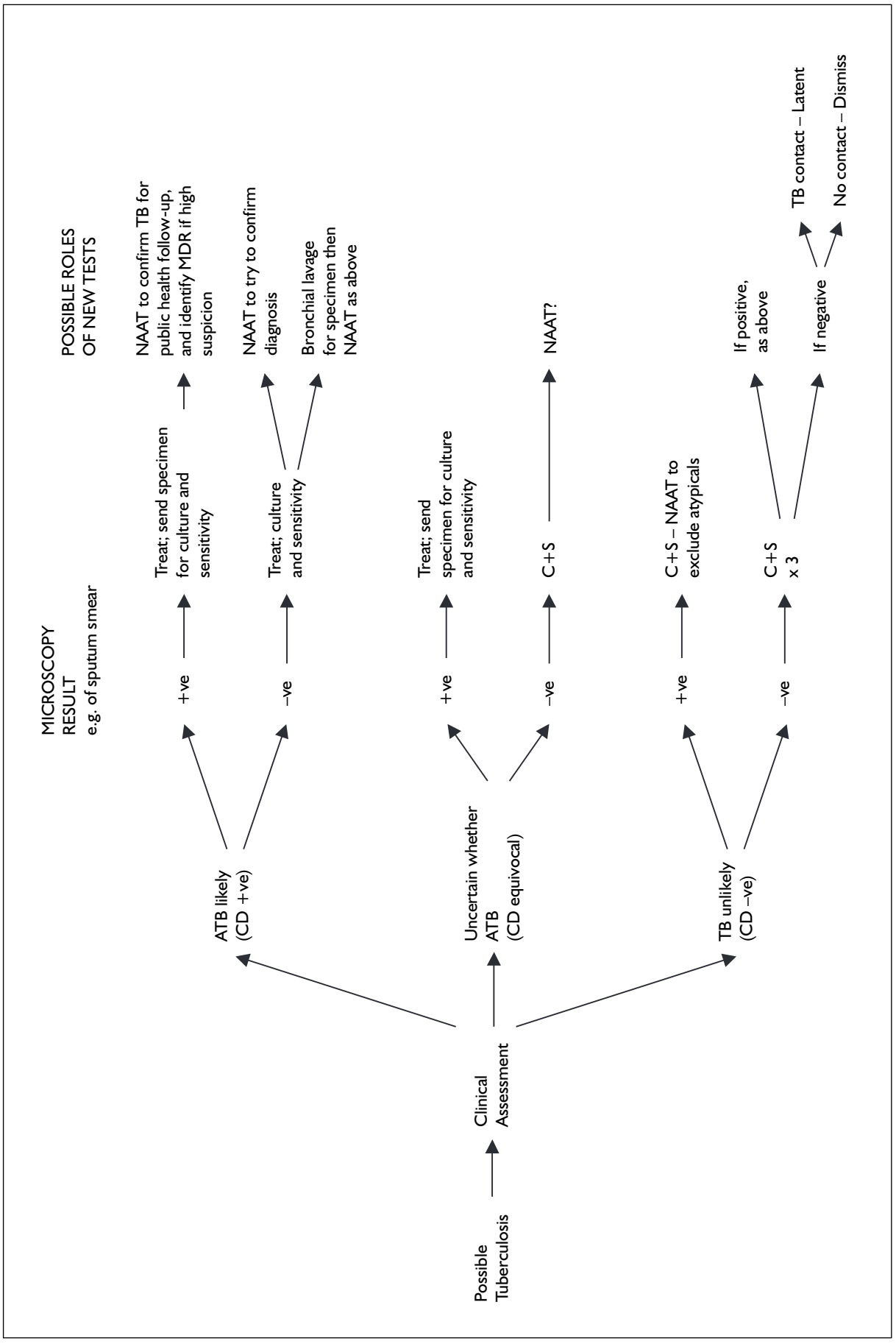


FIGURE 32 *Clinical pathway to diagnosis of suspected ATB*

positive NAAT test will be of any benefit in terms of decision to treat. At the same time, a negative test result in this group would be of little use in ruling out the presence of TB.

Where the degree of clinical suspicion is low or intermediate (around 30% or less), a positive NAAT test should increase clinical certainty of TB to a maximum of 83%. Hence a positive NAAT test is less likely to be a false positive and, taken along with the full clinical picture, could help justify treatment; the NAAT result is just one piece of evidence and would not alone determine the decision. It would be the combination of whatever factors are making the case equivocal plus the NAAT result. One example might be in an HIV-positive individual, where the chest X-ray would not necessarily be characteristic, sputum microscopy is often negative, but (later) culture is positive. On the other hand, a negative NAAT test in the same group (low to intermediate clinical suspicion) could further reduce the probability of TB being present to between 16 and 1%. Some judgement needs to be made as to what would be an acceptable lower level of probability, that is, below which TB could 'acceptably' be assumed not to be present.

The calculations in *Table 80* assume that NAAT test sensitivity and specificity would remain constant regardless of degree of clinical suspicion

TABLE 80 *Impact of a positive or negative NAAT test result on the probability of pulmonary TB in smear-negative patients*

Pre-test probability of TB		Post-test probability of TB[a] NAAT+[b]	Post-test probability of TB[a] NAAT–[b]
High	0.80	*0.98*	0.53
	0.60	*0.95*	0.30
	0.40	0.89	0.16
Intermediate	0.30	0.83	0.11
	0.20	0.74	0.07
	0.10	0.56	0.03
Low	0.08	0.49	*0.02*
	0.05	0.38	*0.01*
	0.03	0.23	*0.01*

[a] Post-test probability of TB following either positive NAAT test result (NAAT+) or negative NAAT test result (NAAT–). Values in italics indicate probability of disease >10% (i.e. treatment on the basis of a positive test result incurs unnecessary treatment in up to 10% of patients), or probability of disease <2% (i.e. ruling out TB on the basis of a negative test result would miss true cases in up to 2% of patients).

[b] NAAT likelihood ratios (LRs) estimated from sensitivity 73.4%, specificity 93.7% (*Table 7*): LR+ = 11.65; LR– = 0.28.

when, in actual fact, sensitivity is likely to be higher for those with high clinical suspicion than for those in whom TB is not considered likely, thereby making the NAAT tests even less useful where clinical suspicion is low.

For extra-pulmonary TB, clinical judgement has both poor sensitivity and specificity. For example, although meningitis may be diagnosed accurately, whether or not it is TB associated is much harder to judge and not infrequently patients are placed on conventional antibacterial therapy, anti-TB therapy and antiviral therapy, initially reflecting this uncertainty. Most studies have focused on improving the speed rather than the accuracy of existing laboratory culture, explaining their frequent focus on culture as the main standard. The NICE guidelines recommend the use of culture, histology and/or chest X-ray for patients with non-respiratory TB.[9]

We found that for pleural TB, ADA and interferon-γ cytokine tests have high sensitivity but limited specificity. More data on the false-positive rate of these tests in representative patients is required, particularly in patients with infected pleural effusions not due to TB. NAATs have high specificity (higher than ADA and interferon-γ) and are therefore likely to confirm the presence of TB infection when positive and could be used alongside these other tests, that is, one could potentially simultaneously use ADA (or interferon-γ) for high sensitivity to rule out disease and NAAT for high specificity to rule it in. *Table 81* presents an example of the impact of applying either the NAAT or ADA test alone, or of sequentially combining them on detection of pleural TB. On their own, neither test really helps to rule in TB, except when clinical suspicion is already very high (60% or over). Given the concerns expressed above regarding clinical diagnosis in extra-pulmonary TB, there will not be many cases in which a positive NAAT or ADA test will change clinical practice. The low sensitivity of the NAAT test in pleural TB also means that a negative NAAT test on its own would not help to rule out pleural TB, unless clinical suspicion was already very low. A negative ADA test, however, could reduce the probability of pleural TB from as much as 20 to 2%. However, the combination of a negative NAAT test and a negative ADA test reduces the probability of pleural TB to 2% or less in most cases and to less than 10% regardless of degree of clinical suspicion. Similarly, if both tests are positive, the likelihood of TB being present is increased to over 90% in those with a pretest probability of disease of 10% or more.

TABLE 81 *Impact of NAAT and ADA tests alone or in combination on the probability of pleural TB*

		Post-test probability of TB[a]							
		NAAT alone[b]		ADA alone[c]		NAAT–		NAAT+	
Pre-test probability of TB		NAAT+	NAAT–	ADA+	ADA–	ADA+	ADA–	ADA+	ADA–
High	0.80	*0.98*	0.54	*0.97*	0.25	*0.91*	0.09	1.00	0.80
	0.60	*0.95*	0.30	*0.93*	0.11	0.79	0.04	0.99	0.60
	0.40	0.88	0.16	0.85	0.05	0.62	*0.02*	*0.98*	0.40
Intermediate	0.30	0.83	0.11	0.78	0.04	0.52	*0.01*	*0.98*	0.30
	0.20	0.74	0.07	0.68	*0.02*	0.38	*0.01*	*0.96*	0.20
	0.10	0.56	0.03	0.49	*0.01*	0.22	0.00	*0.92*	0.10
Low	0.08	0.48	*0.02*	0.41	*0.01*	0.17	0.00	0.89	0.07
	0.05	0.38	*0.02*	0.31	*0.00*	0.12	0.00	0.84	0.05
	0.03	0.23	*0.01*	0.18	*0.00*	0.06	0.00	0.71	*0.02*

[a] Post-test probability of TB following either positive NAAT test result (NAAT+) or negative NAAT test result (NAAT–). Values in italics indicate probability of disease >10% (i.e. treatment on the basis of a positive test result incurs unnecessary treatment in up to 10% of patients), or probability of disease <2% (i.e. ruling out TB on the basis of a negative test result would miss true cases in up to 2% of patients).
[b] NAAT likelihood ratios (LRs) estimated from sensitivity 72.6%, specificity 93.7% (*Table 33*): LR+ = 11.52; LR– = 0.29.
[c] ADA LRs estimated from sensitivity 92.4%; specificity 89.1% (*Table 36*): LR+ = 8.48; LR– = 0.09.

TABLE 82 *Impact of NAAT and ADA tests alone or in combination on the probability of TB meningitis*

		Post-test probability of TB[a]							
		NAAT alone[b]		ADA alone[c]		NAAT–		NAAT+	
Pre-test probability of TB		NAAT+	NAAT–	ADA+	ADA–	ADA+	ADA–	ADA+	ADA–
High	0.80	*0.98*	0.63	*0.97*	0.72	*0.94*	0.20	*1.00*	*0.90*
	0.60	*0.96*	0.39	*0.93*	0.37	*0.85*	0.09	*1.00*	0.78
	0.40	*0.91*	0.22	0.86	0.13	0.72	0.04	*0.99*	0.61
Intermediate	0.30	0.87	0.16	0.80	0.06	0.63	0.03	*0.98*	0.50
	0.20	0.79	0.10	0.69	*0.02*	0.49	*0.02*	*0.97*	0.37
	0.10	0.63	0.05	0.50	*0.01*	0.30	*0.01*	*0.94*	0.20
Low	0.08	0.56	0.03	0.42	*0.00*	0.24	*0.01*	*0.92*	0.16
	0.05	0.45	*0.02*	0.32	*0.00*	0.17	*0.00*	0.88	0.11
	0.03	0.28	*0.01*	0.19	*0.00*	0.09	*0.00*	0.78	0.06

[a] Post-test probability of TB following either positive NAAT test result (NAAT+) or negative NAAT test result (NAAT–). Value in italics indicate probability of disease >10% (i.e. treatment on the basis of a positive test result incurs unnecessary treatment in up to 10% of patients), or probability of disease <2% (i.e. ruling out TB on the basis of a negative test result would miss true cases in up to 2% of patients).
[b] NAAT likelihood ratios (LRs) estimated from sensitivity 58.6%, specificity 96.2% (*Table 51*): LR+ = 15.42; LR– = 0.43.
[c] ADA LRs estimated from sensitivity 86.5%; specificity 90.5% (*Table 53*): LR+ = 9.11; LR– = 0.15.

For TB meningitis, similarly to pleural TB, ADA could be used alongside NAAT to provide high sensitivity and high specificity, respectively. *Table 82* shows that where both tests are positive, the probability of TB meningitis is over 90% in all but the lowest prevalence settings. Where both tests are negative, the probability of TB really being present is 3% or less in those with intermediate to low probability of disease before testing.

The evidence base for ADA in TBM, however, is much smaller than for pleural TB, and more research is needed.

Judicious use of the above rapid tests could significantly impact on patient care and public health by reducing the time elapsed before accurate diagnosis and initiation of treatment, thereby reducing morbidity and mortality (especially in miliary TB and TB meningitis),

reducing the number of healthcare workers and other patients exposed to TB and reducing the number of patients on inappropriate therapy.

Interferon-γ assays for the rapid identification of latent tuberculosis infection

Interferon-γ assays were found to be superior to both TST- and PPD-based assays in all areas of test performance assessed in this review, except for the two studies from a high-prevalence country. In low-prevalence countries, all studies strongly suggest that the RD1-specific antigen-based assays are more accurate than TST for diagnosis of LTBI. If their superior diagnostic capability is found to hold up in routine clinical practice, they will confer several advantages on TB control and prevention programmes by reducing unnecessary chemoprophylaxis, lessening of number of cases with ATB and decreasing unnecessary use of healthcare resources through fewer clinic visits and decreased workload in contact clinics.

The NICE guidelines suggest that interferon-γ testing should be performed in patients with a positive TST or in whom the TST may be unreliable,[9] although of course this will have cost implications for the NHS.

Recommendations for research

Active TB

As a general point, diagnostic accuracy must be established, preferably prospectively, in a wide spectrum of patients, against an appropriate reference test, and avoiding the major sources of bias such as verification bias, lack of blinding and inclusion of all indeterminate results. Large, well-designed primary studies are the key to establishing diagnostic accuracy as a whole and within appropriate subgroups of patients.

- For pulmonary TB, a large, well-designed study of the accuracy of NAAT in clinical diagnosis equivocal smear-negative patients is needed, basically to identify how high a proportion of false-positive results would be generated in this population.
- The place of ADA, interferon-γ and lysozyme for diagnosis of pleural TB each warrants further investigation, and must include patients with non-tuberculous pleural effusions of infective origin to determine the specificity of these tests in routine practice.
- The place of ADA for diagnosis of TB meningitis needs to be established.
- For both pleural and TBM, the combination of NAAT tests with other tests such as ADA should be examined, as their use in combination has the potential to have a huge impact on diagnosis of these diseases.
- The incremental value of combinations of tests, particularly for samples of biological fluids, needs assessment in large, prospective, well-designed studies recruiting representative samples of patients.

Interferon-γ assays for the rapid identification of latent tuberculosis infection

- For LTBI, longitudinal cohort studies to confirm the positive predictive value of interferon-γ assays for subsequent development of ATB should also be performed. Such studies should include a significant proportion of individuals with risk factors for progression to ATB, including young age, HIV infection and iatrogenic immunosuppression.
- For assessing their utility in the diagnostic evaluation of patients with suspected ATB, large prospective trials in routine clinical practice are required. Such studies should include a significant proportion of individuals with factors commonly associated with false-negative TST results, including young age, HIV infection and iatrogenic immunosuppression. These trials should evaluate the performance of the main existing commercial assays (whole blood interferon-γ ELISA and ELISPOT assays) in head-to-head comparisons.
- The role of adding more TB-specific antigens to try to improve diagnostic sensitivity further needs to be assessed.

Acknowledgements

We are very grateful to Liz Payne for carrying out the literature searches, Maxine Caws for assisting with the screening of studies, Christine Clar for data extraction and Ziggy Woodward for acquisition of the papers and managing the Reference Manager database.

Dr Kunst would also like to acknowledge the support of Dr Katherine Fielding and Dr Phillip C Hill from the Department of Epidemiology and Population Health and Dr Alison Grant and Dr Peter Godfrey-Faussett from the Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine.

This work was commissioned by the NHS R&D HTA Programme.

Contribution of authors

This project was funded by the HTA Programme. The grantholders were Jacqueline Dinnes (Senior Research Fellow), Jon Deeks (Senior Medical Statistician), Francis Drobniewski (Director of the Health Protection Agency National Mycobacterium Reference Unit), Ajit Lalvani (Wellcome Senior Clinical Research Fellow and Honorary Consultant Physician) and Norman Waugh (Professor of Public Health). Literature searches for studies in active TB were devised by J Dinnes, and for studies in latent TB by Heinke Kunst (Consultant in Respiratory Medicine). Data extraction for active TB studies was done by J Dinnes and Andrea Gibson (Research Fellow), and for studies of latent TB by H Kunst. Clinical effectiveness analysis was by J Dinnes, J Deeks and H Kunst with comments from F Drobniewksi, A Lalvani and N Waugh. Expert clinical advice was provided throughout by F Drobniewski and A Lalvani. The cost-effectiveness section was by Ewen Cummins (Health Economist) and N Waugh. All authors commented on drafts and the final version.

References

1. Dye C, Scheele S, Dolin P, Pathania V, Raviglione MC. Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. *JAMA* 1999; **282**:677–86.
2. Davies PDO. *Clinical tuberculosis*. London: Arnold; 2003.
3. Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC, *et al*. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch Intern Med* 2003;**163**:1009–21.
4. Nettleman MD, Geerdes H, Roy MC. The cost-effectiveness of preventing tuberculosis in physicians using tuberculin skin testing or a hypothetical vaccine. *Arch Intern Med* 1997; **157**:1121–7.
5. Chemotherapy and management of tuberculosis in the United Kingdom: recommendations 1998. Joint Tuberculosis Committee of the British Thoracic Society. *Thorax* 1998;**53**:536–48.
6. Vynnycky E, Fine PE. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. *Epidemiol Infect* 1997;**119**:183–201.
7. Frieden TR, Sterling TR, Munsiff SS, Watt CJ, Dye C. Tuberculosis. *Lancet* 2003;**362**:887–99.
8. Department of Health. *Stopping tuberculosis in England: an action plan from the Chief Medical Officer.* London: Department of Health; 2004.
9. National Collaborating Centre for Chronic Conditions. *Tuberculosis: clinical diagnosis and management of tuberculosis and measures for its prevention and control*. London: Royal College of Physicians; 2006.
10. Drobniewski FA, Gibson A, Ruddy M, Yates MD. Evaluation and utilization as a public health tool of a national molecular epidemiological tuberculosis outbreak database within the United Kingdom from 1997 to 2001. *J Clin Microbiol* 2003;**41**:1861–8.
11. Rajakumar K, Shafi J, Smith RJ, Stabler RA, Andrew PW, Modha D, *et al*. Use of genome level-informed PCR as a new investigational approach for analysis of outbreak-associated *Mycobacterium tuberculosis* isolates. *J Clin Microbiol* 2004; **42**:1890–6.
12. Stop TB Partnership and World Health Organization. The global plan to stop TB 2006–2015. Geneva: WHO, 2006.
13. Tuberculosis Section, Communicable Diseases Surveillance Centre, Health Protection Agency. Annual Report on Tuberculosis Cases Reported in 2001 in England, Wales and Northern Ireland. March 2004. London: Health Protection Agency.
14. Maguire H, Dale JW, McHugh TD, Butcher PD, Gillespie SH, Costetsos A, *et al*. Molecular epidemiology of tuberculosis in London 1995–7 showing low rate of active transmission. *Thorax* 2002;**57**:617–22.
15. Feleke Y, Abdulkadir J, Aderaye G. Prevalence and clinical features of tuberculosis in Ethiopian diabetic patients. *East Afr Med J* 1999;**76**:361–4.
16. John GT, Shankar V, Abraham AM, Mukundan U, THomas PP, Jacob CK. Risk factors for post-transplant tuberculosis. *Kidney Int* 2001;**60**:1148–53.
17. Centers for Disease Control and Prevention (CDC). Tuberculosis associated with blocking agents against tumor necrosis factor-alpha – California, 2002–2003. *MMWR Morb Mortal Wkly Rep* 2004;**53**:683–6.
18. Shafer RW, Edlin BR. Tuberculosis in patients infected with human immunodeficiency virus: perspective on the past decade. *Clin Infect Dis* 1996;**22**:683–704.
19. Narain JP, Lo YR. Epidemiology of HIV-TB in Asia. *Indian J Med Res* 2004;**120**:277–89.
20. Pelly T, Moore DA, Gilman R, Evans C. Recent tuberculosis advances in Latin America. *Curr Opin Infect Dis* 2004;**17**:397–403.
21. Houk VN, Baker JH, Sorensen K, Kent DC. The epidemiology of tuberculosis infection in a closed environment. *Arch Environ Health* 1968;**16**:26–35.
22. Banner AS. Tuberculosis. Clinical aspects and diagnosis. *Arch Intern Med* 1979;**139**:1387–90.
23. Trinker M, Hofler G, Sill H. False-positive diagnosis of tuberculosis with PCR. *Lancet* 1996; **348**:1388.
24. Verma K, Kapila K. Aspiration cytology for diagnosis of tuberculosis – perspectives in India. *Indian J Pediatr* 2002;**69** Suppl 1:S39–43.
25. van Rie A, Warren R, Richardson M, Victor TC, Gie RP, Enarson DA, *et al*. Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment. *N Engl J Med* 1999;**341**:1174–9.

26. Leiner S, Mays M. Diagnosing latent and active pulmonary tuberculosis: a review for clinicians. *Nurse Pract* 1996;**21**:86, 88, 91–2.

27. Brandli O. The clinical presentation of tuberculosis. *Respiration* 1998;**65**:97–105.

28. Horowitz HW, Luciano BB, Kadel JR, Wormser GP. Tuberculin skin test conversion in hospital employees vaccinated with bacille Calmette–Guérin; recent *Mycobacterium tuberculosis* infection or booster effect. *Am J Infect Control* 1995;**23**:181–8.

29. Comstock GW. False tuberculin test results. *Chest* 1975;**68**(3 Suppl):465–9.

30. Salfinger M, Pfyffer GE. The new diagnostic mycobacteriology laboratory. *Eur J Clin Microbiol Infect Dis* 1994;**13**:961–79.

31. Inderlied CB. *Mycobacteria*. In: Armstrong D, Cohen J, editors. *Infectious diseases*. London: Mosby, Harcourt Publishers; 1999. pp. 22.1–22.20.

32. Lein AD, von Reyn CF. *In vitro* cellular and cytokine responses to mycobacterial antigens: application to diagnosis of tuberculosis infection and assessment of response to mycobacterial vaccines. *Am J Med Sci* 1997;**313**:364–71.

33. Kumar N, Jain S, Murthy NS. Utility of repeat fine needle aspiration in acute suppurative lesions. Follow-up of 263 cases. *Acta Cytol* 2004;**48**:337–40.

34. Yassin MA, Olobo JO, Kidane D, Negesse Y, Shimeles E, Tadesse A, *et al*. Diagnosis of tuberculous lymphadenitis in Butajira, rural Ethiopia. *Scand J Infect Di*s 2003;**35**:240–3.

35. Moore DF, Curry JI. Detection and identification of *Mycobacterium tuberculosis* directly from sputum sediments by Amplicor PCR. *J Clin Microbiol* 1995;**33**:2686–91.

36. Cook VJ, Manfreda J, Hershfield ES. Tuberculous lymphadenitis in Manitoba: incidence, clinical characteristics and treatment. *Can Respir J* 2004; **11**:279–86.

37. Berger HW, Mejia E. Tuberculous pleurisy. *Chest* 1973;**63**:88–92.

38. Ferrer J. Pleural tuberculosis. *Eur Respir J* 1997; **10**:942–7.

39. Maartens G, Bateman ED. Tuberculous pleural effusions: increased culture yield with bedside inoculation of pleural fluid and poor diagnostic value of adenosine deaminase. *Thorax* 1991; **46**:96–9.

40. Seibert AF, Haynes J Jr, Middleton R, Bass JB Jr. Tuberculous pleural effusion. Twenty-year experience. *Chest* 1991;**99**:883–6.

41. Barnes PF. Diagnosing latent tuberculosis infection: the 100-year upgrade. *Am J Respir Crit Care Med* 2001;**163**:807–8.

42. Valdes L, Alvarez D, San Jose E, Penela P, Valle JM, Garcia-Pazos JM, *et al*. Tuberculous pleurisy: a study of 254 patients. *Arch Intern Med* 1998; **158**:2017–21.

43. Hosoglu S, Geyik MF, Balik I, Aygen B, Erol S, Aygencel SG, *et al*. Tuberculous meningitis in adults in Turkey: epidemiology, diagnosis, clinic and laboratory. *Eur J Epidemiol* 2003;**18**:337–43.

44. Behr MA, Warren SA, Salamon H, Hopewell PC, Ponce dL, Daley CL, *et al*. Transmission of *Mycobacterium tuberculosis* from patients smear-negative for acid-fast bacilli. *Lancet* 1999; **353**:444–9.

45. Sutlas PN, Unal A, Forta H, Senol S, Kirbas D. Tuberculous meningitis in adults: review of 61 cases. *Infection* 2003;**31**:387–91.

46. Garg RK. Tuberculosis of the central nervous system. *Postgrad Med J* 1999;**75**:133–40.

47. Thwaites GE, Chau TT, Farrar JJ. Improving the bacteriological diagnosis of tuberculous meningitis. *J Clin Microbiol* 2004;**42**:378–9.

48. Thwaites G, Chau TT, Mai NT, Drobniewski F, McAdam K, Farrar J. Tuberculous meningitis. *J Neurol Neurosurg Psychiatry* 2000;**68**:289–99.

49. Thwaites GE, Chau TT, Stepniewska K, Phu NH, Chuong LV, Sinh DX, *et al*. Diagnosis of adult tuberculous meningitis by use of clinical and laboratory features. *Lancet* 2002;**360**:1287–92.

50. Berenguer J, Moreno S, Laguna F, Vicente T, Adrados M, Ortega A, *et al*. Tuberculous meningitis in patients infected with the human immunodeficiency virus. *N Engl J Med* 1992; **326**:668–72.

51. Whiting P, Rutjes A, Dinnes J, Reitsma J, Bossuyt P, Kleijnen J. Development and validation of methods for assessing the quality and reporting of diagnostic studies. *Health Technol Assess* 2004; **8**(25).

52. Pai M, Riley LW, Colford JM Jr. Interferon-gamma assays in the immunodiagnosis of tuberculosis: a systematic review. *Lancet Infect Dis* 2004;**4**:761–76.

53. Arend SM, Ottenhoff TH, Andersen P, van Dissel JT. Uncommon presentations of tuberculosis: the potential value of a novel diagnostic assay based on the *Mycobacterium tuberculosis*-specific antigens ESAT-6 and CFP-10. *Int J Tuberc Lung Dis* 2001;**5**:680–6.

54. Cegielski JP, Devlin BH, Morris AJ, Kitinya JN, Pulipaka UP, Lema LE, *et al*. Comparison of PCR, culture, and histopathology for diagnosis of tuberculous pericarditis. *J Clin Microbiol* 1997; **35**:3254–7.

55. Demir K, Okten A, Kaymakoglu S, Dincer D, Besisik F, Cevikbas U, *et al*. Tuberculous peritonitis – reports of 26 cases, detailing diagnostic and therapeutic problems. *Eur J Gastroenterol Hepatol* 2001;**13**:581–5.

56. Pai M, Flores LL, Pai N, Hubbard A, Riley LW, Colford JM Jr. Diagnostic accuracy of nucleic acid amplification tests for tuberculous meningitis: a systematic review and meta-analysis.[see comment]. *Lancet Infect Dis* 2003;**3**:633–43.

57. Bernhard JS, Bhatia G, Knauer CM. Gastrointestinal tuberculosis: an eighteen-patient experience and review. *J Clin Gastroenterol* 2000; **30**:397–402.

58. Vazquez ME, Gomez-Cerezo J, Atienza SM, Vazquez Rodriguez JJ. Computed tomography findings of peritoneal tuberculosis: systematic review of seven patients diagnosed in 6 years (1996–2001). *Clin Imaging* 2004;**28**:340–3.

59. Pai M, Flores LL, Hubbard A, Riley LW, Colford JM Jr. Nucleic acid amplification tests in the diagnosis of tuberculous pleuritis: a systematic review and meta-analysis. *BMC Infect Dis* 2004;**4**:6.

60. Enhanced Tuberculosis Surveillance MycobNet (Mycobacterial Surveillance Network). Annual report on tuberculosis cases reported in 2001 in England, Wales and Northern Ireland. URL: http://www.hpa.org.uk/infections/topics_az/tb/pdf/2001_Annual_Report.pdf. 2005.

61. Christensen WI. Genitourinary tuberculosis: review of 102 cases. *Medicine (Baltimore)* 1974;**53**:377–90.

62. Shammaa MZ, Hadidy S, al Asfari R, Siragel-Din MN. Urinary tuberculosis: experience of a teaching hospital in Syria. *Int Urol Nephrol* 1992;**24**:471–80.

63. Kidane D, Olobo JO, Habte A, Negesse Y, Aseffa A, Abate G, *et al*. Identification of the causative organism of tuberculous lymphadenitis in ethiopia by PCR. *J Clin Microbiol* 2002;**40**:4230–4.

64. Campbell IA, Dyson AJ. Lymph node tuberculosis: a comparison of various methods of treatment. *Tubercle* 1977;**58**:171–9.

65. Mortier E, Pouchot J, Girard L, Boussougant Y, Vinceneux P. Assessment of urine analysis for the diagnosis of tuberculosis. *BMJ* 1996;**312**:27–8.

66. Kumar D, Watson JM, Charlett A, Nicholas S, Darbyshire JH. Tuberculosis in England and Wales in 1993: results of a national survey. Public Health Laboratory Service/British Thoracic Society/Department of Health Collaborative Group. *Thorax* 1997;**52**:1060–7.

67. Getachew A, Tesfahunegn Z. Is fine needle aspiration cytology a useful tool for the diagnosis of tuberculous lymphadenitis? *East Afr Med J* 1999; **76**:260–3.

68. Francis IM, Das DK, Luthra UK, Sheikh Z, Sheikh M, Bashir M. Value of radiologically guided fine needle aspiration cytology (FNAC) in the diagnosis of spinal tuberculosis: a study of 29 cases. *Cytopathology* 1999;**10**:390–401.

69. Pertuiset E, Beaudreuil J, Liote F, Horusitzky A, Kemiche F, Richette P, *et al*. Spinal tuberculosis in adults. A study of 103 cases in a developed country, 1980–1994. *Medicine (Baltimore)* 1999; **78**:309–20.

70. Control and prevention of tuberculosis in the United Kingdom: code of practice 2000. Joint Tuberculosis Committee of the British Thoracic Society. *Thorax* 2000;**55**:887–901.

71. Centers for Disease Control and Prevention (CDC); American Thoracic Society. Update: adverse event data and revised American Thoracic Society/CDC recommendations against the use of rifampin and pyrazinamide for treatment of latent tuberculosis infection – United States, 2003. *MMWR Morb Mortal Wkly Rep* 2003;**52**:735–9.

72. Andersen P, Munk ME, Pollock JM, Doherty TM. Specific immune-based diagnosis of tuberculosis. *Lancet* 2000;**356**:1099–104.

73. Gordin F, Slutkin G. The validity of acid-fast smears in the diagnosis of pulmonary tuberculosis. *Arch Pathol Lab Med* 1990;**114**:1025–7.

74. Attorri S, Dunbar S, Clarridge JE, III. Assessment of morphology for rapid presumptive identification of *Mycobacterium tuberculosis* and *Mycobacterium kansasii*. *J Clin Microbiol* 2000; **38**:1426–9.

75. Siddiqi K, Lambert ML, Walley J. Clinical diagnosis of smear-negative pulmonary tuberculosis in low-income countries: the current evidence. *Lancet Infect Dis* 2003;**3**:288–96.

76. Gil-Setas A, Torriba L, Fernandez JL, Martinez-Artola V, Olite J. Evaluation of the MB/BacT system compared with Middlebrook 7h11 and Lowestein–Jensen media for detection and recovery of mycobacteria from clinical specimens. *Clin Microbiol Infect* 2004;**10**:224–8.

77. Hanna BA, Ebrahimzadeh A, Elliott LB, Morgan MA, Novak SM, Rusch-Gerdes S, *et al*. Multicenter evaluation of the BACTEC MGIT 960 system for recovery of mycobacteria. *J Clin Microbiol* 1999;**37**:748–52.

78. Abe C, Hosojima S, Fukasawa Y, Kazumi Y, Takahashi M, Hirano K, *et al*. Comparison of MB-Check, BACTEC, and egg-based media for recovery of mycobacteria. *J Clin Microbiol* 1992; **30**:878–81.

79. Welch DF, Guruswamy AP, Sides SJ, Shaw CH, Gilchrist MJ. Timely culture for mycobacteria which utilizes a microcolony method. *J Clin Microbiol* 1993;**31**:2178–84.

80. Watterson SA, Drobniewski FA. Modern laboratory diagnosis of mycobacterial infections. *J Clin Pathol* 2000;**53**:727–32.

81. Bothamley GH. Serological diagnosis of tuberculosis. *Eur Respir J Suppl* 1995;**20**:676s–88s.

82. Gounder C, Queiroz Mello FC, Conde MB, Bishai WR, Kritski AL, Chaisson RE, *et al*. Field evaluation of a rapid immunochromatographic test for tuberculosis. *J Clin Microbiol* 2002;**40**:1989–93.

83. Al Zahrani K, Al Jahdali H, Poirier L, Rene P, Gennaro ML, Menzies D. Accuracy and utility of commercially available amplification and serologic tests for the diagnosis of minimal pulmonary tuberculosis. *Am J Respir Crit Care Med* 2000; **162**(4 Pt 1):1323–9.

84. Lyashchenko K, Colangeli R, Houde M, Al Jahdali H, Menzies D, Gennaro ML. Heterogeneous antibody responses in tuberculosis. *Infect Immun* 1998;**66**:3936–40.

85. Pottumarthy S, Wells VC, Morris AJ. A comparison of seven tests for serological diagnosis of tuberculosis. *J Clin Microbiol* 2000;**38**:2227–31.

86. Greco S, Girardi E, Masciangelo R, Capoccetta GB, Saltini C. Adenosine deaminase and interferon gamma measurements for the diagnosis of tuberculous pleurisy: a meta-analysis. *Int J Tuberc Lung Dis* 2003;**7**:777–86.

87. Riantawan P, Chaowalit P, Wongsangiem M, Rojanaraweewong P. Diagnostic value of pleural fluid adenosine deaminase in tuberculous pleuritis with reference to HIV coinfection and a Bayesian analysis. *Chest* 1999;**116**:97–103.

88. Morgan MA, Horstmeier CD, DeYoung DR, Roberts GD. Comparison of a radiometric method (BACTEC) and conventional culture media for recovery of mycobacteria from smear-negative specimens. *J Clin Microbiol* 1983;**18**:384–8.

89. Tortoli E, Cichero P, Piersimoni C, Simonetti MT, Gesu G, Nista D. Use of BACTEC MGIT 960 for recovery of mycobacteria from clinical specimens: multicenter study. *J Clin Microbiol* 1999;**37**:3578–82.

90. Sharp SE, Lemes M, Erlich SS, Poppiti RJ Jr. A comparison of the Bactec 9000MB system and the Septi-Chek AFB system for the detection of mycobacteria. *Diagn Microbiol Infect Dis* 1997; **28**:69–74.

91. Cruciani M, Scarparo C, Malena M, Bosco O, Serpelloni G, Mengoli C. Meta-analysis of BACTEC MGIT 960 and BACTEC 460 TB, with or without solid media, for detection of mycobacteria. *J Clin Microbiol* 2004;**42**:2321–5.

92. Drobniewski FA, Caws M, Gibson A, Young D. Modern laboratory diagnosis of tuberculosis. *Lancet Infect Dis* 2003;**3**:141–7.

93. Harvell JD, Hadley WK, Ng VL. Increased sensitivity of the BACTEC 460 mycobacterial radiometric broth culture system does not decrease the number of respiratory specimens required for a definitive diagnosis of pulmonary tuberculosis. *J Clin Microbiol* 2000;**38**:3608–11.

94. Kaminski DA, Hardy DJ. Selective utilization of DNA probes for identification of *Mycobacterium* species on the basis of cord formation in primary BACTEC 12B cultures. *J Clin Microbiol* 1995; **33**:1548–50.

95. Diaz-Infantes MS, Ruiz-Serrano MJ, Martinez-Sanchez L, Ortega A, Bouza E. Evaluation of the MB/BacT mycobacterium detection system for susceptibility testing of *Mycobacterium tuberculosis*. *J Clin Microbiol* 2000;**38**:1988–9.

96. Goloubeva V, Lecocq M, Lassowsky P, Matthys F, Portaels F, Bastian I. Evaluation of mycobacteria growth indicator tube for direct and indirect drug susceptibility testing of *Mycobacterium tuberculosis* from respiratory specimens in a Siberian prison hospital. *J Clin Microbiol* 2001;**39**:1501–5.

97. Tortoli E, Mattei R, Savarino A, Bartolini L, Beer J. Comparison of *Mycobacterium tuberculosis* susceptibility testing performed with BACTEC 460TB (Becton Dickinson) and MB/BacT (Organon Teknika) systems. *Diagn Microbiol Infect Dis* 2000;**38**:83–6.

98. Tortoli E, Benedetti M, Fontanelli A, Simonetti MT. Evaluation of automated BACTEC MGIT 960 system for testing susceptibility of *Mycobacterium tuberculosis* to four major antituberculous drugs: comparison with the radiometric BACTEC 460TB method and the agar plate method of proportion. *J Clin Microbiol* 2002;**40**:607–10.

99. Bemer P, Palicova F, Rusch-Gerdes S, Drugeon HB, Pfyffer GE. Multicenter evaluation of fully automated BACTEC Mycobacteria Growth Indicator Tube 960 system for susceptibility testing of *Mycobacterium tuberculosis*. *J Clin Microbiol* 2002;**40**:150–4.

100. Banaiee N, Bobadilla-Del-Valle M, Bardarov SJ, Riska PF, Small PM, Ponce-De-Leon A, *et al*. Luciferase reporter mycobacteriophages for detection, identification, and antibiotic susceptibility testing of *Mycobacterium tuberculosis* in Mexico. *J Clin Microbiol* 2001;**39:**3883–8.

101. Louie M, Louie L, Simor AE. The role of DNA amplification technology in the diagnosis of infectious diseases. *CMAJ* 2000;**163**:301–9.

102. American Thoracic Society Workshop. Rapid diagnostic tests for tuberculosis. What is the appropriate use? *Am J Respir Crit Care Med* 1997; **155**:1804–14.

103. Centers for Disease Control and Prevention. Nucleic acid amplification tests for tuberculosis. *MMWR Morb Mortal Wkly Rep* 1996;**45**:950–2.

104. McHugh TD, Pope CF, Ling CL, Patel S, Billington OJ, Gosling RD, *et al*. Prospective evaluation of BDProbeTec strand displacement amplification (SDA) system for diagnosis of tuberculosis in non-respiratory and respiratory samples. *J Med Microbiol* 2004;**53**(Pt 12):1215–19.

105. Kent L, McHugh TD, Billington O, Dale JW, Gillespie SH. Demonstration of homology between IS6110 of *Mycobacterium tuberculosis* and DNAs of other *Mycobacterium* spp. *J Clin Microbiol* 1995; **33**:2290–3.

106. El Dawi TG, Saeed eN, Hamid ME. Evaluation of a PCR-amplified IS6110 insertion element in the rapid diagnosis of pulmonary tuberculosis in comparison to microscopic methods in Sudan. *Saudi Med J* 2004;**25**:1644–7.

107. Kivihya-Ndugga LE, van Cleeff MR, Githui WA, Nganga LW, Kibuga DK, Odhiambo JA, *et al*. A comprehensive comparison of Ziehl–Neelsen and fluorescence microscopy for the diagnosis of tuberculosis in a resource-poor urban setting. *Int J Tuberc Lung Dis* 2003;**7**:1163–71.

108. Jatana SK, Nair MN, Lahiri KK, Sarin NP. Polymerase chain reaction in the diagnosis of tuberculosis. *Indian Pediatr* 2000;**37**:375–82.

109. Barnes PF. Rapid diagnostic tests for tuberculosis: progress but no gold standard. *Am J Respir Crit Care Med* 1997;**155**:1497–8.

110. Jouveshomme S, Cambau E, Trystram D, Szpytma M, Sougakoff W, Derenne JP, *et al*. Clinical utility of an amplification test based on ligase chain reaction in pulmonary tuberculosis. *Am J Respir Crit Care Med* 1998;**158**:1096–101.

111. Fadda G, Ardito F, Sanguinetti M, Posteraro B, Ortona L, Chezzi C, *et al*. Evaluation of the Abbott LCx *Mycobacterium tuberculosis* assay in comparison with culture methods in selected Italian patients. *New Microbiol* 1998;**21**:97–103.

112. Eltringham IJ, Wilson SM, Drobniewski FA. Evaluation of a bacteriophage-based assay (phage amplified biologically assay) as a rapid screen for resistance to isoniazid, ethambutol, streptomycin, pyrazinamide, and ciprofloxacin among clinical isolates of Mycobacterium tuberculosis. *J Clin Microbiol* 1999;**37**:3528–32.

113. Wilson SM, al Suwaidi Z, McNerney R, Porter J, Drobniewski F. Evaluation of a new rapid bacteriophage-based method for the drug susceptibility testing of *Mycobacterium tuberculosis*. *Nat Med* 1997;**3**:465–8.

114. Riska PF, Su Y, Bardarov S, Freundlich L, Sarkis G, Hatfull G, *et al*. Rapid film-based determination of antibiotic susceptibilities of *Mycobacterium tuberculosis* strains by using a luciferase reporter phage and the Bronx Box. *J Clin Microbiol* 1999; **37**:1144–9.

115. Butler WR, Guthertz LS. Mycolic acid analysis by high-performance liquid chromatography for identification of *Mycobacterium* species. *Clin Microbiol Rev* 2001;**14**:704–26, table.

116. Eisenstadt J, Hall GS. Microbiology and classification of mycobacteria. *Clin Dermatol* 1995; **13**:197–206.

117. Anargyros P, Astill DS, Lim IS. Comparison of improved BACTEC and Lowenstein–Jensen media for culture of mycobacteria from clinical specimens. *J Clin Microbiol* 1990;**28**:1288–91.

118. Hale YM, Desmond EP, Jost KC Jr, Salfinger M. Access to newer laboratory procedures: a call for action. *Int J Tuberc Lung Dis* 2000;**4**(12 Suppl 2): S171–5.

119. Scarparo C, Piccoli P, Rigon A, Ruggiero G, Nista D, Piersimoni C. Direct identification of mycobacteria from MB/BacT alert 3D bottles: comparative evaluation of two commercial probe assays. *J Clin Microbiol* 2001;**39**:3222–7.

120. Alcaide F, Benitez MA, Escriba JM, Martin R. Evaluation of the BACTEC MGIT 960 and the MB/BacT systems for recovery of mycobacteria from clinical specimens and for species identification by DNA AccuProbe. *J Clin Microbiol* 2000;**38**:398–401.

121. Lebrun L, Espinasse F, Poveda JD, Vincent-Levy-Frebault V. Evaluation of nonradioactive DNA probes for identification of mycobacteria. *J Clin Microbiol* 1992;**30**:2476–8.

122. Bartfai Z, Somoskovi A, Kodmon C, Szabo N, Puskas E, Kosztolanyi, L, *et al*. Molecular characterization of rifampin-resistant isolates of *Mycobacterium tuberculosis* from Hungary by DNA sequencing and the line probe assay. *J Clin Microbiol* 2001;**39**:3736–9.

123. Suffys PN, da Silva RA, de Oliveira M, Campos CE, Barreto AM, Portaels F, *et al*. Rapid identification of *Mycobacteria* to the species level using INNO-LiPA Mycobacteria, a reverse hybridization assay. *J Clin Microbiol* 2001;**39**:4477–82.

124. Tortoli E, Nanetti A, Piersimoni C, Cichero P, Farina C, Mucignat G, *et al*. Performance assessment of new multiplex probe assay for identification of mycobacteria. *J Clin Microbiol* 2001;**39**:1079–84.

125. Trombert-Paolantoni S, Poveda JD, Figarella P. Comparison of two reverse hybridization methods for mycobacterial identification in clinical practice. *Pathol Biol (Paris)* 2004;**52**:462–8.

126. Yam WC, Tam CM, Leung CC, Tong HL, Chan KH, Leung ET, *et al*. Direct detection of rifampin-resistant *Mycobacterium tuberculosis* in respiratory specimens by PCR–DNA sequencing. *J Clin Microbiol* 2004;**42**:4438–43.

127. Drobniewski FA, Watterson SA, Wilson SM, Harris GS. A clinical, microbiological and economic analysis of a national service for the rapid molecular diagnosis of tuberculosis and rifampicin resistance in *Mycobacterium tuberculosis*. *J Med Microbiol* 2000;**49**:271–8.

128. Telenti A, Imboden P, Marchesi F, Lowrie D, Cole S, Colston MJ, *et al*. Detection of rifampicin-

resistance mutations in *Mycobacterium tuberculosis*. *Lancet* 1993;**341**:647–50.

129. Torres MJ, Criado A, Palomares JC, Aznar J. Use of real-time PCR and fluorimetry for rapid detection of rifampin and isoniazid resistance-associated mutations in *Mycobacterium tuberculosis*. *J Clin Microbiol* 2000;**38**:3194–9.

130. Leao SC, Bernardelli A, Cataldi A, Zumarraga M, Robledo J, Realpe T, *et al*. Multicenter evaluation of mycobacteria identification by PCR restriction enzyme analysis in laboratories from Latin America and the Caribbean. *J Microbiol Methods* 2005;**61**:193–9.

131. da Silva CF, Ueki SY, Geiger DC, Leao SC. hsp65 PCR-restriction enzyme analysis (PRA) for identification of mycobacteria in the clinical laboratory. *Rev Inst Med Trop Sao Paulo* 2001; **43**:25–8.

132. Wong DA, Yip PC, Tse DL, Tung VW, Cheung DT, Kam KM. Routine use of a simple low-cost genotypic assay for the identification of mycobacteria in a high throughput laboratory. *Diagn Microbiol Infect Dis* 2003;**47**:421–6.

133. Huang TS, Liu YC, Huang WK, Chen CS, Tu HZ, Cheng DL. Evaluation of polymerase chain reaction-restriction enzyme analysis of mycobacteria cultured in BACTEC 12B bottles. *J Formos Med Assoc* 1996;**95**:530–5.

134. Drobniewski FA, Watt B, Smith EG, Magee JG, Williams R, Holder J, *et al*. A national audit of the laboratory diagnosis of tuberculosis and other mycobacterial diseases within the United Kingdom 5461. *J Clin Pathol* 1999;**52**:334–7.

135. Chaisson RE, Keruly JC, McAvinue S, Gallant JE, Moore RD. Effects of an incentive and education program on return rates for PPD test reading in patients with HIV infection. *J Acquir Immune Defic Syndr Hum Retrovirol* 1996;**11**:455–9.

136. Nardell EA, Fan D, Shepard JA, Mark EJ. Case records of the Massachusetts General Hospital. Weekly clinicopathological exercises. Case 22-2004. A 30-year-old woman with a pericardial effusion. *N Engl J Med* 2004**;351**:279–87.

137. Stuart RL, Bennett N, Forbes A, Grayson ML. A paired comparison of tuberculin skin test results in health care workers using 5 TU and 10 TU tuberculin. *Thorax* 2000;**55**:693–5.

138. Targeted tuberculin testing and treatment of latent tuberculosis infection. American Thoracic Society. *MMWR Recomm Rep* 2000;**49**(RR-6):1–51.

139. Pouchot J, Grasland A, Collet C, Coste J, Esdaile JM, Vinceneux P. Reliability of tuberculin skin test measurement. *Ann Intern Med* 1997;**126**:210–14.

140. Bearman JE, Kleinman H, Glyer VV, Lacroix OM. A study of variability in tuberculin test reading. *Am Rev Respir Dis* 1964;**90**:913–19.

141. Fine PE, Bruce J, Ponnighaus JM, Nkhosa P, Harawa A, Vynnycky E. Tuberculin sensitivity: conversions and reversions in a rural African population. *Int J Tuberc Lung Dis* 1999;**3**:962–75.

142. Floyd S, Ponnighaus JM, Bliss L, Nkhosa P, Sichali L, Msiska G, *et al*. Kinetics of delayed-type hypersensitivity to tuberculin induced by bacille Calmette–Guérin vaccination in northern Malawi. *J Infect Dis* 2002;**186**:807–14.

143. Huebner RE, Schein MF, Bass JB Jr. The tuberculin skin test. *Clin Infect Dis* 1993;**17**:968–75.

144. Kwamanga DO, Swai OB, Agwanda R, Githui W. Effect of non-tuberculous *Mycobacteria* infection on tuberculin results among primary school children in Kenya. *East Afr Med J* 1995;**72**:222–7.

145. 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.

146. von Reyn CF, Green PA, McCormick D, Huitt GA, Marsh BJ, Magnusson M, *et al*. Dual skin testing with *Mycobacterium avium* sensitin and purified protein derivative: an open study of patients with *M. avium* complex infection or tuberculosis. *Clin Infect Dis* 1994;**19**:15–20.

147. von Reyn CF, Williams DE, Horsburgh CR Jr, Jaeger AS, Marsh BJ, Haslov K, *et al*. Dual skin testing with *Mycobacterium avium* sensitin and purified protein derivative to discriminate pulmonary disease due to *M. avium* complex from pulmonary disease due to *Mycobacterium tuberculosis*. *J Infect Dis* 1998;**177**:730–6.

148. Menzies D. Interpretation of repeated tuberculin tests. Boosting, conversion, and reversion. *Am J Respir Crit Care Med* 1999;**159**:15–21.

149. Chin DP, Osmond D, Page-Shafer K, Glassroth J, Rosen MJ, Reichman LB, *et al*. Reliability of anergy skin testing in persons with HIV infection. The Pulmonary Complications of HIV Infection Study Group. *Am J Respir Crit Care Med* 1996; **153**(6 Pt 1):1982–4.

150. Cellestis. Clinicians Guide to Quantiferon-TB Gold. URL: www.cellestis.com/IRM/contentAU/gold/cliniciansguide.pdf. 2004.

151. Oxford Immunotec Limited. T Spot TB. URL: http://www.oxfordimmunotec.com/products.htm. 2004.

152. Mabtech AB. ELISpot – assay procedure. URL: http://www.mabtech.com/elispot.asp. 2004.

153. Chapman AL, Munkanta M, Wilkinson KA, Pathan AA, Ewer K, Ayles H, *et al*. Rapid detection of active and latent tuberculosis infection in HIV-positive individuals by enumeration of Mycobacterium tuberculosis-specific T cells. *AIDS* 2002;**16**:2285–93.

154. Fietta A, Meloni F, Cascina A, Morosini M, Marena C, Troupioti P, *et al*. Comparison of a whole-blood interferon-gamma assay and tuberculin skin testing in patients with active tuberculosis and individuals at high or low risk of *Mycobacterium tuberculosis* infection. *Am J Infect Control* 2003;**31**:347–53.

155. Pottumarthy S, Morris AJ, Harrison AC, Wells VC. Evaluation of the tuberculin gamma interferon assay: potential to replace the Mantoux skin test. *J Clin Microbiol* 1999;**37**:3229–32.

156. Streeton JA, Desem N, Jones SL. Sensitivity and specificity of a gamma interferon blood test for tuberculosis infection. *Int J Tuberc Lung Dis* 1998; **2**:443–50.

157. Desem N, Jones SL. Development of a human gamma interferon enzyme immunoassay and comparison with tuberculin skin testing for detection of Mycobacterium tuberculosis infection. *Clin Diagn Lab Immunol* 1998;**5**:531–6.

158. Mazurek GH, LoBue PA, Daley CL, Bernardo J, Lardizabal AA, Bishai WR, *et al*. Comparison of a whole-blood interferon gamma assay with tuberculin skin testing for detecting latent *Mycobacterium tuberculosis* infection. *JAMA* 2001; **286**:1740–7.

159. Mazurek GH, Villarino ME. Guidelines for using the QuantiFERON-TB test for diagnosing latent *Mycobacterium tuberculosis* infection. Centers for Disease Control and Prevention. *MMWR Recomm Rep* 2003;**52**(RR-2):15–18.

160. Brock I, Munk ME, Kok-Jensen A, Andersen P. Performance of whole blood IFN-gamma test for tuberculosis diagnosis based on PPD or the specific antigens ESAT-6 and CFP-10. *Int J Tuberc Lung Dis* 2001;**5**:462–7.

161. van Pinxteren LA, Ravn P, Agger EM, Pollock J, Andersen P. Diagnosis of tuberculosis based on the two specific antigens ESAT-6 and CFP10. *Clin Diagn Lab Immunol* 2000;**7**:155–60.

162. Lalvani A, Nagvenkar P, Udwadia Z, Pathan AA, Wilkinson KA, Shastri JS, *et al*. Enumeration of T cells specific for RD1-encoded antigens suggests a high prevalence of latent *Mycobacterium tuberculosis* infection in healthy urban Indians. *J Infect Dis* 2001;**183**:469–77.

163. Harboe M, Oettinger T, Wiker HG, Rosenkrands I, Andersen P. Evidence for occurrence of the ESAT-6 protein in Mycobacterium tuberculosis and virulent *Mycobacterium bovis* and for its absence in *Mycobacterium bovis* BCG. *Infect Immun* 1996; **64**:16–22.

164. Pollock JM, Andersen P. The potential of the ESAT-6 antigen secreted by virulent mycobacteria for specific diagnosis of tuberculosis. *J Infect Dis* 1997;**175**:1251–4.

165. Liu XQ, Dosanjh D, Varia H, Ewer K, Cockle P, Pasvol G, *et al*. Evaluation of T-cell responses to novel RD1- and RD2-encoded *Mycobacterium tuberculosis* gene products for specific detection of human tuberculosis infection. *Infect Immun* 2004; **72**:2574–81.

166. Lalvani A, Pathan AA, Durkan H, Wilkinson KA, Whelan A, Deeks JJ, *et al*. Enhanced contact tracing and spatial tracking of *Mycobacterium tuberculosis* infection by enumeration of antigen-specific T cells. *Lancet* 2001;**357**:2017–21.

167. Berthet FX, Rasmussen PB, Rosenkrands I, Andersen P, Gicquel B. A *Mycobacterium tuberculosis* operon encoding ESAT-6 and a novel low-molecular-mass culture filtrate protein (CFP-10). *Microbiology* 1998;**144**(Pt 11):3195–203.

168. Liebeschuetz S, Bamber S, Ewer K, Deeks J, Pathan AA, Lalvani A. Diagnosis of tuberculosis in South African children with a T-cell-based assay: a prospective cohort study. *Lancet* 2004; **364**:2196–203.

169. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA* 2002;**274**:645–51.

170. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JHP, *et al*. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;**282**:1061–6.

171. Rutjes A, Reitsma J, Di Nisio M, Smidt N, Zwinderman AH, Rijn JC, *et.al*. Bias in diagnostic accuracy studies due to shortcomings in design and conduct. Presented at the Xth Annual Cochrane Colloquium, Barcelona, 2003.

172. Small PM, Perkins DP. More rigour needed in trials of new diagnostic agents for tuberculosis. *Lancet* 2000;**356**:1048–9.

173. Heffner JE. Evaluating diagnostic tests in the pleural space. Differentiating transudates from exudates as a model. *Clin Chest Med* 1998; **19**:277–93.

174. Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. In Egger M, Smith GD, Altman DA, editors. *Systematic reviews in health care*. London: BMJ Books; 2001. pp. 248–82.

175. Condos R, McClune A, Rom WN, Schluger NW. Peripheral-blood-based PCR assay to identify patients with active pulmonary tuberculosis. *Lancet* 1996;**347**:1082–5.

176. Lienhardt C, Fielding K, Sillah J, Tunkara A, Donkor S, Manneh K, *et al*. Risk factors for tuberculosis infection in sub-Saharan Africa: a contact study in The Gambia. *Am J Respir Crit Care Med* 2003;**168**:448–55.

177. Richeldi L, Ewer K, Losi M, Bergamini BM, Roversi P, Deeks J, *et al*. T cell-based tracking of multidrug resistant tuberculosis infection following brief exposure. *Am J Respir Crit Care Med* 2004;**170**:288–95.

178. Ewer K, Deeks J, Alvarez L, Bryant G, Waller S, Andersen P, *et al*. Comparison of T-cell-based assay with tuberculin skin test for diagnosis of *Mycobacterium tuberculosis* infection in a school tuberculosis outbreak. *Lancet* 2003;**361**:1168–73.

179. Haley CE, McDonald RC, Rossi L, Jones WD Jr, Haley RW, Luby JP. Tuberculosis epidemic among hospital personnel. *Infect Control Hosp Epidemiol* 1989;**10**:204–10.

180. Hutton MD, Stead WW, Cauthen GM, Bloch AB, Ewing WM. Nosocomial transmission of tuberculosis associated with a draining abscess. *J Infect Dis* 1990;**161**:286–95.

181. Begg CB. Biases in the assessment of diagnostic tests. *Stat Med* 1987;**6**:411–23.

182. Mol BW, Bossuyt PMM. Evaluating the effectiveness of diagnostic tests. PhD Academic Medical Center, University of Amsterdam. Tubal subfertility and ectopic pregnancy: evaluating the effectiveness of diagnostic tests, 1999.

183. Fletcher RH, Fletcher SW, Wagner EH. Studying cases. *Clinical epidemiology: the essentials*. London: Williams & Wilkins; 1996. pp. 208–27.

184. Knottnerus JA, Leffers P. The influence of referral patterns on the characteristics of diagnostic tests. *J Clin Epidemiol* 1992;**45**:1143–54.

185. Miller WC. Bias in discrepant analysis: when two wrongs don't make a right. *J Clin Epidemiol* 1998; **51**:219–31.

186. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, *et al*. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;**282**:1061–6.

187. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data analytic approaches and some additional considerations. *Stat Med* 1993;**12**:1293–316.

188. Sarmiento OL, Weigle KA, Alexander J, Weber DJ, Miller WC. Assessment by meta-analysis of PCR for diagnosis of smear-negative pulmonary tuberculosis. *J Clin Microbiol* 2003;**41**:3233–40.

189. Goto M, Noguchi Y, Koyama H, Hira K, Shimbo T, Fukui T. Diagnostic value of adenosine deaminase in tuberculous pleural effusion: a meta-analysis. *Ann Clin Biochem* 2003;**40**(Pt 4):374–81.

190. Amer S. Usefulness of PCR in the detection of *Mycobacterium tuberculosis* in clinically suspected cases of pulmonary tuberculosis. *J Med Res Inst* 2000;**21**:32–47.

191. Chedore P, Jamieson FB. Routine use of the Gen-Probe MTD2 amplification test for detection of *Mycobacterium tuberculosis* in clinical specimens in a large public health mycobacteriology laboratory. *Diagn Microbiol Infect Dis* 1999;**35**:185–91.

192. Mitarai S, Kurashima A, Tamura A, Nagai H, Shishido H. Clinical evaluation of Amplicor Mycobacterium detection system for the diagnosis of pulmonary mycobacterial infection using sputum. *Tuberculosis (Edinb)* 2001;**81**:319–25.

193. Lockman S, Hone N, Kenyon TA, Mwasekaga M, Villauthapillai M, Creek T, *et al*. Etiology of pulmonary infections in predominantly HIV-infected adults with suspected tuberculosis, Botswana. *Int J Tuberc Lung Dis* 2003;**7**:714–23.

194. Rolfs A, Beige J, Finckh U, Kohler B, Schaberg T, Lokies J, *et al*. Amplification of *Mycobacterium tuberculosis* from peripheral blood. *J Clin Microbiol* 1995;**33**:3312–14.

195. Thierry D, Chureau C, Aznar C, Guesdon JL. The detection of *Mycobacterium tuberculosis* in uncultured clinical specimens using the polymerase chain reaction and a non-radioactive DNA probe. *Mol Cell Probes* 1992;**6**:181–91.

196. Albert H, Heydenrych A, Brookes R, Mole RJ, Harley B, Subotsky E, *et al*. Performance of a rapid phage-based test, FASTPlaqueTB, to diagnose pulmonary tuberculosis from sputum specimens in South Africa. *Int J Tuberc Lung Dis* 2002;**6**:529–37.

197. Muzaffar R, Batool S, Aziz F, Naqvi A, Rizvi A. Evaluation of the FASTPlaqueTB assay for direct detection of *Mycobacterium tuberculosis* in sputum specimens [comment]. *Int J Tuberc Lung Dis* 2002; **6**:635–40.

198. Cavusoglu C, Guneri S, Suntur M, Bilgic A. Clinical evaluation of the FASTPlaqueTB for the rapid diagnosis of pulmonary tuberculosis. *Turk J Med Sci* 2002;**32**:487–92.

199. Albert H, Trollip AP, Mole RJ, Hatch SJ, Blumberg L. Rapid indication of multidrug-resistant tuberculosis from liquid cultures using FASTPlaqueTB-RIF, a manual phage-based test. *Int J Tuberc Lung Dis* 2002;**6**(6):523–8.

200. Conde MB, Marinho SR, Pereira MF, Lapa e Silva JR, Saad MH, Sales CL, *et al*. The usefulness of serum adenosine deaminase 2 (ADA2) activity in adults for the diagnosis of pulmonary tuberculosis. *Respir Med* 2002;**96**(8):607–10.

201. Cho S-N, Lee J-H, Lee H-Y, Won H-J, Chong Y, Chang J, *et al*. Detection of *Mycobacterium tuberculosis* antigens in sputum for the diagnosis of pulmonary tuberculosis. *J Korean Soc Microbiol* 1997;**32**:285–91.

202. Chan CH, Chan RC, Arnold M, Cheung H, Cheung SW, Cheng AF. Bronchoscopy and tuberculostearic acid assay in the diagnosis of

sputum smear-negative pulmonary tuberculosis: a prospective study with the addition of transbronchial biopsy 3297. *Q J Med* 1992;**82**:15–23.

203. Piersimoni C, Scarparo C, Piccoli P, Rigon A, Ruggiero G, Nista D, *et al*. Performance assessment of two commercial amplification assays for direct detection of *Mycobacterium tuberculosis* complex from respiratory and extrapulmonary specimens. *J Clin Microbiol* 2002;**40**:4138–42.

204. Marttila HJ, Soini H, Vyshnevskaya E, Vyshnevskiy BI, Otten TF, Vasilyef AV, *et al*. Line probe assay in the rapid detection of rifampin-resistant *Mycobacterium tuberculosis* directly from clinical specimens. *Scand J Infect Dis* 1999; **31**:269–73.

205. Makhlouf NA, Nassar G, Makhlouf M, Habbal Z. Adenosine deaminase activity in various pathological effusions. *J Med Liban* 1992;**40**:142–4.

206. Ghelani DR, Parikh FS, Hakim AS, Pai-Dhungat JV. Diagnostic significance of immunoglobulins and adenosine deaminase in pleural effusion. *J Assoc Physicians India* 1999;**47**:787–90.

207. Kunter E, Cerrahoglu K, Ilvan A, Isitmangil T, Turken O, Okutan O, *et al*. The value of pleural fluid anti-A60 IgM in BCG-vaccinated tuberculous pleurisy patients. *Clin Microbiol Infect* 2003; **9**:212–20.

208. Chierakul N, Damrongchokpipat P, Chaiprasert A, Arjratanakul W. Antibody detection for the diagnosis of tuberculous pleuritis. *Int J Tuberc Lung Dis* 2001;**5**:968–72.

209. Dhand R, Ganguly NK, Vaishnavi C, Gilhotra R, Malik SK. False-positive reactions with enzyme-linked immunosorbent assay of *Mycobacterium tuberculosis* antigens in pleural fluid. *J Med Microbiol* 1988;**26**:241–3.

210. Chiang CS, Chiang CD, Lin JW, Huang PL, Chu JJ. Neopterin, soluble interleukin-2 receptor and adenosine deaminase levels in pleural effusions. *Respiration* 1994;**61**:150–4.

211. Ribera E, Espanol T, Martinez-Vazquez JM, Ocana I, Encabo G. Lymphocyte proliferation and gamma-interferon production after "in vitro" stimulation with PPD. Differences between tuberculous and nontuberculous pleurisy in patients with positive tuberculin skin test. *Chest* 1990;**97**:1381–5.

212. Nagesh BS, Sehgal S, Jindal SK, Arora SK. Evaluation of polymerase chain reaction for detection of *Mycobacterium tuberculosis* in pleural fluid. *Chest* 2001;**119**:1737–41.

213. Villegas MV, Labrada LA, Saravia NG. Evaluation of polymerase chain reaction, adenosine deaminase, and interferon-gamma in pleural fluid for the differential diagnosis of pleural tuberculosis. *Chest* 2000;**118**:1355–64.

214. Eintracht S, Silber E, Sonnenberg P, Koornhof HJ, Saffer D. Analysis of adenosine deaminase isoenzyme-2 (ADA(2)) in cerebrospinal fluid in the diagnosis of tuberculosis meningitis. *J Neurol Neurosurg Psychiatry* 2000;**69**:137–8.

215. Prabhakar S, Oommen A. ELISA using mycobacterial antigens as a diagnostic aid for tuberculous meningitis. *J Neurol Sci* 1987; **78**:203–11.

216. Wiggelinkhuizen J, Mann M. The radioactive bromide partition test in the diagnosis of tuberculous meningitis in children. *J Pediatr* 1980;**97**:843–7.

217. Baek CH, Kim SI, Ko YH, Chu KC. Polymerase chain reaction detection of *Mycobacterium tuberculosis* from fine-needle aspirate for the diagnosis of cervical tuberculous lymphadenitis. *Laryngoscope* 2000;**110**:30–4.

218. Goel MM, Ranjan V, Dhole TN, Srivastava AN, Mehrotra A, Kushwaha MR, *et al*. Polymerase chain reaction vs. conventional diagnosis in fine needle aspirates of tuberculous lymph nodes. *Acta Cytol* 2001;**45**:333–40.

219. Brisson-Noel A, Gicquel B, Lecossier D, Levy-Frebault V, Nassif X, Hance AJ. Rapid diagnosis of tuberculosis by amplification of mycobacterial DNA in clinical samples. *Lancet* 1989;**2**:1069–71.

220. Kiran U, Shriniwas, Rohatgi M. Laboratory diagnosis of tuberculous lymphadenitis using soluble antigen fluorescent antibody test. *Indian J Med Res* 1982;**76**:1–4.

221. Jain A, Verma RK, Tiwari V, Goel MM. Development of a new antigen detection dot-ELISA for diagnosis of tubercular lymphadenitis in fine needle aspirates. *J Microbiol Methods* 2003; **53**:107–12.

222. Ceyhan M, Kanra G, Secmeer G, Erdem G, Ecevit Z, Yilmaz E, *et al*. Diagnosis of childhood tuberculosis by polymerase chain reaction. *Turk J Pediatr* 1996;**38**:399–405.

223. Brant CQ, Silva MRJ, Macedo EP, Vasconcelos C, Tamaki N, Ferraz ML. The value of adenosine deaminase (ADA) determination in the diagnosis of tuberculous ascites. *Rev Inst Med Trop Sao Paulo* 1995;**37**:449–53.

224. Ribera E, Martinez-Vasquez JM, Ocana I, Ruiz I, Jiminez JG, Encabo G, *et al*. Diagnostic value of ascites gamma interferon levels in tuberculous peritonitis. Comparison with adenosine deaminase activity. *Tubercle* 1991;**72**:193–7.

225. Kaur A, Basha A, Ranjan M, Oommen A. Poor diagnostic value of adenosine deaminase in pleural, peritoneal and cerebrospinal fluids in tuberculosis. *Indian J Med Res* 1992;**95**:270–7.

226. Lee JH, Lee CW, Lee SG, Yang HS, Hong MK, Kim JJ, *et al*. Comparison of polymerase chain

reaction with adenosine deaminase activity in pericardial fluid for the diagnosis of tuberculous pericarditis. *Am J Med* 2002;**113**:519–21.

227. Burgess LJ, Reuter H, Carstens ME, Taljaard JJ, Doubell AF. The use of adenosine deaminase and interferon-gamma as diagnostic tools for tuberculous pericarditis. *Chest* 2002;**122**:900–5.

228. Moussa OM, Eraky I, El Far MA, Osman HG, Ghoneim MA. Rapid diagnosis of genitourinary tuberculosis by polymerase chain reaction and non-radioactive DNA hybridization. *J Urol* 2000; **164**:584–8.

229. Hemal AK, Gupta NP, Rajeev TP, Kumar R, Dar L, Seth P. Polymerase chain reaction in clinically suspected genitourinary tuberculosis: comparison with intravenous urography, bladder biopsy, and urine acid fast bacilli culture. *Urology* 2000; **56**:570–4.

230. Rattan A, Gupta SK, Singh S, Takker D, Kumar S, Bai P, *et al*. Detection of antigens of *Mycobacterium tuberculosis* in patients of infertility by monoclonal antibody based sandwiched enzyme linked immunosorbent assay (ELISA). *Tubere Lung Dis* 1993;**74**:200–3.

231. Van der Spoel van Dijk A, MCleod A, Botha PL, Shipley J, Kapnoudhis M, Beukes CA. The diagnosis of skeletal tuberculosis by polymerase chain reaction. *Cent Afr J Med* 2000;**46**:144–9.

232. Idigoras P, Beristain X, Iturzaeta A, Vicente D, Perez-Trallero E. Comparison of the automated nonradiometric Bactec MGIT 960 system with Lowenstein–Jensen, Coletsos, and Middlebrook 7H11 solid media for recovery of mycobacteria. *Eur J Clin Microbiol Infect Di*s 2000;**19**:350–4.

233. Bachmann LM, Coray R, Estermann P, ter Riet G. Identifying diagnostic studies in MEDLINE: reducing the number needed to read. *J Am Med Inform Assoc* 2002;**9**:653–8.

234. Bachmann LM, Estermann P, Kronenberg C, ter Riet G. Identifying diagnostic accuracy studies in EMBASE. *J Med Libr Assoc* 2003;**91**:341–6.

235. Haynes RB, Wilczynski N, McKibbon KA, Walker CJ, Sinclair JC. Developing optimal search strategies for detecting clinically sound studies in MEDLINE. *J Am Med Inform Assoc* 1994;**1**:447–58.

236. Centre for Reviews and Dissemination. *Undertaking systematic reviews of research on effectiveness. CRD's guidance for carrying out or commissioning reviews*. CRD Report 4, 2nd ed. York: University of York, Centre for Reviews and Dissemination; 2001.

237. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994;**271**:703–7.

238. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1994;**271**:389–91.

239. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol* 1995;**48**:119–30.

240. Whiting P, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;**3**:25.

241. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, *et al*. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000;**283**:2008–12.

242. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, *et al*. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003;**326**:41–4.

243. Egger M, Davey Smith G, Schneider M. Systematic reviews of observational studies. In Egger M, Davey Smith G, Altman D, editors. *Systematic reviews in health care: meta-analysis in context*. London: BMJ Books; 2002. pp. 211–27.

244. Irwig L, Tostesen ANA, Gatsonis C, Lau J, Colditz G, Chalmers TC, *et al*. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med* 1994;**120**:667–76.

245. The Cochrane Collaboration. *Review Manager (RevMan 4.2).* [4.1]. The Cochrane Collaboration; 2003.

246. Deeks JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001;**323**:157–62.

247. Deeks JJ, Higgins JPT, Altman DG. Analysing and presenting results. In Alderson P, Higgins JPT, Altman DG, editors. *Cochrane Reviewers' Handbook 4.2.2 [updated November 2004]*. URL: http://www.cochrane.org/resources/handbook/hbook.htm. Accessed April 2005.

248. WRAIR Study. A comparison of QuantiFERON-TB interferon-gamma test with the TST for detection of *M. tuberculosis* infection in military recruits. In *FDA document on safety and effectiveness of QuantiFERON*. URL: http://www.fda.gov/cdrh/pdf/p010033b.doc. Accessed 1 March 2005.

249. Pathan AA, Wilkinson KA, Klenerman P, McShane H, Davidson RN, Pasvol G, *et al*. Direct *ex vivo* analysis of antigen-specific IFN-gamma-secreting CD4 T cells in *Mycobacterium tuberculosis*-infected individuals: associations with clinical disease state and effect of treatment. *J Immunol* 2001;**167**:5217–25.

250. Ravn P, Munk ME, Andersen AB, Lundgren B, Nielsen LN, Lillebaek T, *et al*. Reactivation of tuberculosis during immunosuppressive treatment in a patient with a positive QuantiFERON-RD1 test. *Scand J Infect Dis* 2004;**36**:499–501.

251. Rolinck-Werninghaus C, Magdorf K, Stark K, Lyashchenko K, Gennaro ML, Colangeli R, *et al*. The potential of recombinant antigens ESAT-6, MPT63 and mig for specific discrimination of *Mycobacterium tuberculosis* and *M. avium* infection. *Eur J Pediatr* 2003;**162**:534–6.

252. Scarpellini P, Tasca S, Galli L, Beretta A, Lazzarin A, Fortis C. Selected pool of peptides from ESAT-6 and CFP-10 proteins for detection of *Mycobacterium tuberculosis* infection. *J Clin Microbiol* 2004;**42**:3469–74.

253. Wu-Hsieh BA, Chen CK, Chang JH, Lai SY, Wu CH, Cheng WC, *et al*. Long-lived immune response to early secretory antigenic target 6 in individuals who had recovered from tuberculosis. *Clin Infect Dis* 2001;**33**:1336–40.

254. Kunst, H. Interferon-gamma assay tests for latent tuberculosis: a systematic review of the literature. Master's thesis. London School of Hygiene and Tropical Medicine; 2004.

255. Bellete B, Coberly J, Barnes GL, Ko C, Chaisson RE, Comstock GW, *et al*. Evaluation of a whole-blood interferon-gamma release assay for the detection of Mycobacterium tuberculosis infection in 2 study populations. *Clin Infect Dis* 2002;**34**:1449–56.

256. Brock I, Weldingh K, Lillebaek T, Follmann F, Andersen P. Comparison of a new specific blood test and the skin test in tuberculosis contacts. *Am J Respir Crit Care Med* 2004;**170**:65–9.

257. Converse PJ, Jones SL, Astemborski J, Vlahov D, Graham NM. Comparison of a tuberculin interferon-gamma assay with the tuberculin skin test in high-risk adults: effect of human immunodeficiency virus infection. *J Infect Dis* 1997;**176**:144–50.

258. Hill PC, Brookes RH, Fox A, Fielding K, Jeffries DJ, Jackson-Sillah D, *et al*. Large-scale evaluation of enzyme-linked immunospot assay and skin test for diagnosis of *Mycobacterium tuberculosis* infection against a gradient of exposure in The Gambia. *Clin Infect Dis* 2004;**38**:966–73.

259. Kimura M, Converse PJ, Astemborski J, Rothel JS, Vlahov D, Comstock GW, *et al*. Comparison between a whole blood interferon-gamma release assay and tuberculin skin testing for the detection of tuberculosis infection among patients at risk for tuberculosis exposure. *J Infect Dis* 1999; **179**:1297–300.

260. Vekemans J, Lienhardt C, Sillah JS, Wheeler JG, Lahai GP, Doherty MT, *et al*. Tuberculosis contacts but not patients have higher gamma interferon responses to ESAT-6 than do community controls in The Gambia. *Infect Immun* 2001;**69**:6554–7.

261. Hill PC, Jackson-Sillah D, Fox A, Franken KL, Lugos MD, Jeffries DJ, *et al*. ESAT-6/CFP-10 fusion protein and peptides for optimal diagnosis of mycobacterium tuberculosis infection by *ex vivo* enzyme-linked immunospot assay in the Gambia. *J Clin Microbiol* 2005;**43**:2070–4.

262. Arend SM, van Meijgaarden KE, de Boer K, De Palou EC, van Soolingen D, Ottenhoff TH, *et al*. Tuberculin skin testing and *in vitro* T cell responses to ESAT-6 and culture filtrate protein 10 after infection with *Mycobacterium marinum* or *M. kansasii*. *J Infect Dis* 2002;**186**:1797–807.

263. Greinert U, Schlaak M, Rusch-Gerdes S, Flad HD, Ernst M. Low *in vitro* production of interferon-gamma and tumor necrosis factor-alpha in HIV-seronegative patients with pulmonary disease caused by nontuberculous mycobacteria. *J Clin Immunol* 2000;**20**:445–52.

264. Geluk A, van Meijgaarden KE, Franken KL, Subronto YW, Wieles B, Arend SM, *et al*. Identification and characterization of the ESAT-6 homologue of *Mycobacterium leprae* and T-cell cross-reactivity with *Mycobacterium tuberculosis*. *Infect Immun* 2002;**70**:2544–8.

265. Johnson PD, Stuart RL, Grayson ML, Olden D, Clancy A, Ravn P, *et al*. Tuberculin-purified protein derivative-, MPT-64-, and ESAT-6-stimulated gamma interferon responses in medical students before and after *Mycobacterium bovis* BCG vaccination and in patients with tuberculosis. *Clin Diagn Lab Immunol* 1999;**6**:934–7.

266. Mori T, Sakatani M, Yamagishi F, Takashima T, Kawabe Y, Nagao K, *et al*. Specific detection of tuberculosis infection: an interferon-gamma-based assay using new antigens. *Am J Respir Crit Care Med* 2004;**170**:59–64.

267. Baily GV. Tuberculosis prevention trial, Madras. *Indian J Med Res* 1980;**72** Suppl:1–74.

268. Eason RJ. Tuberculin sensitivity. *Ann Trop Paediatr* 1987;**7**:87–90.

269. Karalliedde S, Katugaha LP, Uragoda CG. Tuberculin response of Sri Lankan children after BCG vaccination at birth. *Tubercle* 1987;**68**:33–8.

270. Selwyn PA, Sckell BM, Alcabes P, Friedland GH, Klein RS, Schoenbaum EE. High risk of active tuberculosis in HIV-infected drug users with cutaneous anergy. *JAMA* 1992;**268**:504–9.

271. Sodhi A, Gong J, Silva C, Qian D, Barnes PF. Clinical correlates of interferon gamma production in patients with tuberculosis. *Clin Infect Dis* 1997;**25**:617–20.

272. Richeldi L, Ewer K, Losi M, Hansell DM, Roversi P, Fabbri LM, *et al*. Early diagnosis of subclinical multidrug-resistant tuberculosis. *Ann Intern Med* 2004;**140**:709–13.

273. Sester M, Sester U, Clauer P, Heine G, Mack U, Moll T, *et al*. Tuberculin skin testing underestimates a high prevalence of latent tuberculosis infection in hemodialysis patients. *Kidney Int* 2004; **65**:1826–34.

274. Roos BR, van Cleeff MR, Githui WA, Kivihya-Ndugga L, Odhiambo JA, Kibuga DK, *et al*. Cost-effectiveness of the polymerase chain reaction versus smear examination for the diagnosis of tuberculosis in Kenya: a theoretical model. *Int J Tuberc Lung Dis* 1998;**2**:235–41.

275. Steele BA, Daniel TM. Evaluation of the potential role of serodiagnosis of tuberculosis in a clinic in Bolivia by decision analysis. *Am Rev Respir Dis* 1991;**143**:713–16.

276. Lim TK, Cherian J, Poh KL, Leong TY. The rapid diagnosis of smear-negative pulmonary tuberculosis: a cost-effectiveness analysis. *Respirology* 2000;**5**:403–9.

277. Dowdy DW, Maters A, Parrish N, Beyrer C, Dorman SE. Cost-effectiveness analysis of the Gen-Probe amplified *Mycobacterium tuberculosis* direct test as used routinely on smear-positive respiratory specimens. *J Clin Microbiol* 2003; **41**(3):948–53.

278. Heymann SJ, Brewer TF, Ettling M. Effectiveness and cost of rapid and conventional laboratory methods for *Mycobacterium tuberculosis* screening. *Public Health Rep* 1997;**112**:513–23.

279. Rajalahti I, Vuorinen P, Jarvenpaa R, Nieminen MM. *Mycobacterium tuberculosis* complex is not detected by DNA amplification assay in sputum specimens of patients with lung scars due to past pulmonary tuberculosis. *Int J Tuberc Lung Dis* 2003;**7**:190–3.

280. Conaty SJ, Claxton AP, Enoch DA, Hayward AC, Lipman MC, Gillespie SH. The interpretation of nucleic acid amplification tests for tuberculosis: do rapid tests change treatment decisions? *J Infect* 2005;**50**:187–92.

281. Rose DN. Benefits of screening for latent *Mycobacterium tuberculosis* infection. *Arch Intern Med* 2000;**160**:1513–21.

282. Jasmer RM, Nahid P, Hopewell PC. Latent tuberculosis infection. *N Engl J Med* 2002; **347**:1860–6.

283. Song F, Eastwood A, Gilbody S, Duley D, Sutton AJ. Publication and related biases. *Health Technol Assess* 2000;**4**(10).

284. Balasegaram S, Watson JM, Rose AM, Charlett A, Nunn AJ, Rushdy A, *et al*. For the Public Health Laboratory Service/British Thoracic Society/Department of Health Collaborative Group. A decade of change: tuberculosis in England and Wales 1988–98. *Arch Dis Child* 2003;**88**:772–7.

285. Drobniewski FA. Diagnosing multidrug resistant tuberculosis in Britain. *BMJ* 1998;**317**:1263–4.

286. Deville W, Yzermans N, Bouter LM, Bezemer PD, van der Windt DA. Heterogeneity in systematic reviews of diagnostic studies. In *Proceedings of the 2nd Symposium on Systematic Reviews: Beyond the Basics*, Oxford, 1999. Conference abstract. Available from: www.mrw.interscience.wiley.com/cochrane/clcmr/articles/CMR-983/frame.html

287. Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, *et al*. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med* 1994; **120**:667–76.

288. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol* 1995;**48**:119–30.

289. Abe C, Hirano K, Wada M, Kazumi Y, Takahashi M, Fukasawa Y, *et al*. Detection of *Mycobacterium tuberculosis* in clinical specimens by polymerase chain reaction and Gen-Probe Amplified Mycobacterium Tuberculosis Direct Test. *J Clin Microbiol* 1993;**31**:3270–4.

290. Abu-Amero KK. Potential for the use of polymerase chain reaction (PCR) in the detection and identification of *Mycobacterium tuberculosis* complex in sputum samples. *Mol Biol Today* 2002;**3**:39–42.

291. Alcala L, Ruiz-Serrano MJ, Hernangomez S, Marin M, de Viedma DG, San Juan R, *et al*. Evaluation of the upgraded amplified *Mycobacterium tuberculosis* direct test (Gen-Probe) for direct detection of *Mycobacterium tuberculosis* in respiratory and non-respiratory specimens. *Diagn Microbiol Infect Dis* 2001;**41**:51–6.

292. Alonso P, Orduna A, Bratos MA, San Miguel A, Rodriguez-Torres A. Clinical evaluation of a commercial ligase-based gene amplification method for detection of *Mycobacterium tuberculosis*. *Eur J Clin Microbiol Infect Dis* 1998;**17**:371–6.

293. Arimura M, Ohuchi T, Suzuki Y, Hishinuma A, Oikawa S, Sato J, *et al*. Clinical significance of direct detection of *Mycobacterium tuberculosis* in respiratory specimens by polymerase chain reaction. *Dokkyo J Med Sci* 1996;**22**:143–8.

294. Bemer-Melchior P, Boudigueux V, Drugeon HB. Clinical validity of an automated DNA amplification system for diagnosis of pulmonary tuberculosis. *Medicines et Maladies Infectieuses* 2000;**30**:253–61.

295. Bennedsen J, Thomsen VO, Pfyffer GE, Funke G, Feldmann K, Beneke A, *et al*. Utility of PCR in diagnosing pulmonary tuberculosis. *J Clin Microbiol* 1996;**34**:1407–11.

296. Bergmann JS, Woods GL. Clinical evaluation of the Roche AMPLICOR PCR *Mycobacterium tuberculosis* test for detection of *M. tuberculosis* in respiratory specimens. *J Clin Microbiol* 1996; **34**:1083–5.

297. Bergmann JS, Yuoh G, Fish G, Woods GL. Clinical evaluation of the enhanced Gen-Probe Amplified *Mycobaterium tuberculosis* Direct Test for rapid diagnosis of tuberculosis in prison inmates. *J Clin Microbiol* 1999;**37**:1419–25.

298. Cartuyvels R, De Ridder C, Jonckheere S, Verbist L, Van Eldere J. Prospective clinical evaluation of Amplicor *Mycobacterium tuberculosis* PCR test as a screening method in a low-prevalence population. *J Clin Microbiol* 1996;**34**:2001–3.

299. Catanzaro A, Perry S, Clarridge JE, Dunbar S, Goodnight-White S, LoBue PA, *et al*. The role of clinical suspicion in evaluating a new diagnostic test for active tuberculosis: results of a multicenter prospective trial. *JAMA* 2000;**283**:639–45.

300. Chin DP, Yajko DM, Hadley WK, Sanders CA, Nassos PS, Madej JJ, *et al*. Clinical utility of a commercial test based on the polymerase chain reaction for detecting *Mycobacterium tuberculosis* in respiratory specimens. *Am J Respir Crit Care Med* 1995;**151**:1872–7.

301. Cohen RA, Muzaffar S, Schwartz D, Bashir S, Luke S, McGartland LP, *et al*. Diagnosis of pulmonary tuberculosis using PCR assays on sputum collected within 24 hours of hospital admission. *Am J Respir Crit Care Med* 1998; **157**:156–61.

302. D'Amato RF, Wallman AA, Hochstein LH, Colaninno PM, Scardamaglia M, Ardila E, *et al*. Rapid diagnosis of pulmonary tuberculosis by using Roche AMPLICOR *Mycobacterium tuberculosis* PCR test. *J Clin Microbiol* 1995;**33**:1832–4.

303. Denis O, Devaster JM, Vandenberg O, Vanachter H, Lafontaine T, Lin C, *et al*. Evaluation of ligase chain reaction for direct detection of *Mycobacterium tuberculosis* in respiratory specimens. *Zentralbl Bakteriol* 1998;**288**:59–65.

304. Devallois A, Legrand E, Rastogi N. Evaluation of Amplicor MTB test as adjunct to smears and culture for direct detection of Mycobacterium tuberculosis in the French Caribbean. *J Clin Microbiol* 1996;**34**:1065–8.

305. dos Anjos Filho L, Oelemann W, Barreto CE, Kritski AL, de Souza Fonseca L. Sensitivity of AMPLICOR MTB on direct detection of *Mycobacterium tuberculosis* in smear-negative specimens from outpatients in Rio de Janeiro. *Braz J Microbiol* 2002;**33**:163–5.

306. Ehlers S, Ignatius R, Regnath T, Hahn H. Diagnosis of extrapulmonary tuberculosis by Gen-Probe amplified *Mycobacterium tuberculosis* direct test. *J Clin Microbiol* 1996;**34**:2275–9.

307. Eing BR, Becker A, Sohns A, Ringelmann R. Comparison of Roche Cobas Amplicor *Mycobacterium tuberculosis* assay with in-house PCR and culture for detection of *M. tuberculosis*. *J Clin Microbiol* 1998;**36**:2023–9.

308. Gleason Beavis K, Lichty MB, Jungkind DL, Giger O. Evaluation of Amplicor PCR for direct detection of *Mycobacterium tuberculosis* from sputum specimens. *J Clin Microbiol* 1995;**33**:2582–6.

309. Gomez-Pastrana D, Torronteras R, Caro P, Anguita ML, Lopez-Barrio AM, Andres A, *et al*. Comparison of Amplicor, in-house polymerase chain reaction, and conventional culture for the diagnosis of tuberculosis in children. *Clin Infect Dis* 2001;**32**:17–22.

310. Hoffner SE, Norberg R, Carlos-Toro J, Winqvist N, Koivula T, Dias F, *et al*. Direct detection of *Mycobacterium tuberculosis* in sputum samples from Guinea Bissau by an rRNA target-amplified test system. *Tuber Lung Dis* 1996;**77**:67–70.

311. Hoffner SE, Cristea M, Klintz L, Petrini B, Kallenius G. RNA amplification for direct detection of *Mycobacterium tuberculosis* in respiratory samples. *Scand J Infect Dis* 1996; **28**:59–61.

312. Kambashi B, Mbulo G, McNerney R, Tembwe R, Kambashi A, Tihon V, *et al*. Utility of nucleic acid amplification techniques for the diagnosis of pulmonary tuberculosis in sub-Saharan Africa. *Int J Tuberc Lung Dis* 2001;**5**:364–9.

313. Kang EY, Choi JA, Seo BK, Oh YW, Lee CK, Shim JJ. Utility of polymerase chain reaction for detecting *Mycobacterium tuberculosis* in specimens from percutaneous transthoracic needle aspiration. *Radiology* 2002;**225**:205–9.

314. La Rocco MT, Wanger A, Ocera H, Macias E. Evaluation of a commercial rRNA amplification assay for direct detection of *Mycobacterium tuberculosis* in processed sputum. *Eur J Clin Microbiol Infect Dis* 1994;**13**:726–31.

315. Lim TK, Gough A, Chin NK, Kumarasinghe G. Relationship between estimated pretest probability and accuracy of automated *Mycobacterium tuberculosis* assay in smear-negative pulmonary tuberculosis. *Chest* 2000;**118**:641–7.

316. Lim TK, Zhu D, Gough A, Lee KH, Kumarasinghe G. What is the optimal approach for using a direct amplification test in the routine diagnosis of pulmonary tuberculosis? A preliminary assessment. *Respirology* 2002;**7**:351–7.

317. Lindbrathen A, Gaustad P, Hovig B, Tonjum T. Direct detection of *Mycobacterium tuberculosis complex* in clinical samples from patients in Norway by ligase chain reaction. *J Clin Microbiol* 1997; **35**:3248–53.

318. Middleton AM, Cullinan P, Wilson R, Kerr JR, Chadwick MV. Interpreting the results of the amplified *Mycobacterium tuberculosis* direct test for detection of *M. tuberculosis* rRNA. *J Clin Microbiol* 2003;**41**:2741–3.

319. Mitarai S, Tanoue S, Sugita C, Sugihara E, Tamura A, Nagono Y, *et al*. Potential use of

Amplicor PCR kit in diagnosing pulmonary tuberculosis from gastric aspirate. *J Microbiol Methods* 2001;**47**:339–44.

320. Neu N. Diagnosis of pediatric tuberculosis in the modern era. *Pediatric Infect Dis J* 1999;**18**:122–6.

321. Osumi M, Toyoda T, Kawashiro T, Aoyagi T. Detection of *Mycobacterium tuberculosis* in clinical specimens other than sputum by a specific DNA probe with amplification of the ribosomal RNA. *Kansenshogaku Zasshi* 1995;**69**:1376–82.

322. Piersimoni C, Zitti P, Cimarelli ME, Nista D, De Sio G. Clinical utility of the Gen-Probe amplified *Mycobacterium tuberculosis* direct test compared with smear and culture for the diagnosis of pulmonary tuberculosis. *Clin Microbiol Infect* 1998;**4**:442–6.

323. Reischl U, Lehn N, Wolf H, Naumann L. Clinical evaluation of the automated COBAS AMPLICOR MTB assay for testing respiratory and nonrespiratory specimens. *J Clin Microbiol* 1998; **36**:2853–60.

324. Sato K, Tomioka H, Kawahara S, Shishido S. Evaluation of two commercial diagnostic kits for *Mycobacterium tuberculosis* completely based on bacterial DNA and rRNA amplification for direct detection of tubercle bacilli in sputum specimens. *Kansenshogaku Zasshi* 1998;**72**:504–11.

325. Se Thoe SY, Tay L, Sng EH. Evaluation of Amplicor- and IS6110-PCR for direct detection of *Mycobacterium tuberculosis complex* in Singapore. *Trop Med Int Health* 1997;**2**:1095–101.

326. Shim TS, Chi HS, Lee SD, Koh Y, Kim WS, Kim DS, *et al*. Adequately washed bronchoscope does not induce false-positive amplification tests on bronchial aspirates in the diagnosis of pulmonary tuberculosis. *Chest* 2002;**121**:774–81.

327. Smith MB, Bergmann JS, Onoroto M, Mathews G, Woods GL. Evaluation of the enhanced amplified *Mycobacterium tuberculosis* direct test for direct detection of *Mycobacterium tuberculosis* complex in respiratory specimens. *Arch Pathol Lab Med* 1999; **123**:1101–13.

328. Viinanen AH, Soini H, Marjamaki M, Liippo K, Viljanen MK. Ligase chain reaction assay is clinically useful in the discrimination of smear-positive pulmonary tuberculosis from atypical mycobacterioses. *Ann Med* 2000;**32**:279–83.

329. Vuorinen P, Miettinen A, Vuento R, Hallstrom O. Direct detection of *Mycobacterium tuberculosis complex* in respiratory specimens by Gen-Probe Amplified Mycobacterium Tuberculosis Direct Test and Roche Amplicor Mycobacterium Tuberculosis Test. *J Clin Microbiol* 1995;**33**:1856–9.

330. Wang SX, Tay L. Evaluation of three nucleic acid amplification methods for direct detection of *Mycobacterium tuberculosis complex* in respiratory specimens. *J Clin Microbiol* 1999;**37**:1932–4.

331. Yam WC, Yuen KY, Seto WH. Direct detection of *Mycobacterium tuberculosis* in respiratory specimens using an automated DNA amplification assay and a single tube nested polymerase chain reaction (PCR). *Clin Chem Lab Med* 1998;**36**:597–9.

332. Yee YC, Gough A, Kumarasinghe G, Lim TK. The pattern of utilisation and accuracy of a commercial nucleic acid amplification test for the rapid diagnosis of *Mycobacterium tuberculosis* in routine clinical practice. *Singapore Med J* 2002;**43**:415–20.

333. Zambardi G, Druetta A, Roure C, Fouque B, Girardo P, Chypre C, *et al*. Rapid diagnosis of *Mycobacterium tuberculosis* infections by an ELISA-like detection of polymerase chain reaction products. *Mol Cell Probes* 1995;**9**:91–9.

334. Afghani B, Stutman HR. Polymerase chain reaction for diagnosis of *M. tuberculosis*: comparison of simple boiling and a conventional method for DNA extraction. *Biochem Mol Med* 1996;**57**:14–18.

335. Albay A, Kisa O, Baylan O, Doganci L. The evaluation of FASTPlaqueTB test for the rapid diagnosis of tuberculosis. *Diagn Microbiol Infect Dis* 2003;**46**:211–15.

336. Alfonso R, Romero RE, Patarroyo ME, Murillo LA. Mtp-40 and alpha antigen gene fragment amplification for the detection of *Mycobacterium tuberculosis* in Colombian clinical specimens. *Mem Inst Oswaldo Cruz* 2002;**97**:1157–63.

337. Al Ghamdi B, Karawya E. The utility of polymerase chain reaction in the diagnosis of pulmonary tuberculosis in smear-negative, culture-negative patients. *Bahrain Med Bull* 1998;**20**:12–13.

338. Bahrmand AR, Bakayev VV, Babaei MH. Use of polymerase chain reaction for primary diagnosis of pulmonary tuberculosis in the clinical laboratory. *Scand J Infect Dis* 1996;**28**:469–72.

339. Beige J, Lokies J, Schaberg T, Finckh U, Fischer M, Mauch H, *et al*. Clinical evaluation of a *Mycobacterium tuberculosis* PCR assay. *J Clin Microbiol* 1995;**33**:90–5.

340. Choi YJ, Hu Y, Mahmood A. Clinical significance of a polymerase chain reaction assay for the detection of *Mycobacterium tuberculosis*. *Am J Clin Pathol* 1996;**105**:200–4.

341. Delacourt C, Poveda JD, Chureau C, Beydon N, Mahut B, de Blic J, *et al*. Use of polymerase chain reaction for improved diagnosis of tuberculosis in children. *J Pediatr* 1995;**126**(5 Pt 1):703–9.

342. Eisenach KD, Sifford MD, Cave MD, Bates JH, Crawford JT. Detection of *Mycobacterium tuberculosis* in sputum samples using a polymerase chain reaction. *Am Rev Respir Dis* 1991;**144**:1160–3.

343. Fauville-Dufaux M, Vanfleteren B, de Wit L, Vincke JP, van Vooren JP, Yates MD, *et al*. Rapid detection of tuberculous and non-tuberculous

mycobacteria by polymerase chain reaction amplification of a 162 bp DNA fragment from antigen 85. *Eur J Clin Microbiol Infect Dis* 1992; **11**:797–803.

344. Ginesu F, Pirina P, Sechi LA, Molicotti P, Santoru L, Porcu L, *et al*. Microbiological diagnosis of tuberculosis: a comparison of old and new methods. *J Chemother* 1998;**10**:295–300.

345. Gomez-Pastrana D, Torronteras R, Caro P, Anguita ML, Barrio AM, Andres A, *et al*. Diagnosis of tuberculosis in children using a polymerase chain reaction. *Pediatr Pulmonol* 1999;**28**:344–51.

346. Gori A, Franzetti F, Marchetti G, Catozzi L, Corbellino M. Specific detection of *Mycobacterium tuberculosis* by mtp40 nested PCR. *J Clin Microbiol* 1996;**34**:2866–7.

347. Herrera EA, Segovia M. Evaluation of mtp40 genomic fragment amplification for specific detection of *Mycobacterium tuberculosis* in clinical specimens. *J Clin Microbiol* 1996;**34**:1108–13.

348. Kocagoz T, Yilmaz E, Ozkara S, Kocagoz S, Hayran M, Sachedeva M, *et al*. Detection of *Mycobacterium tuberculosis* in sputum samples by polymerase chain reaction using a simplified procedure. *J Clin Microbiol* 1993;**31**:1435–8.

349. Kolk AH, Schuitema AR, Kuijper S, van Leeuwen J, Hermans PW, van Embden JD, *et al*. Detection of *Mycobacterium tuberculosis* in clinical samples by using polymerase chain reaction and a nonradioactive detection system. *J Clin Microbiol* 1992;**30**:2567–75.

350. Kox LF, Rhienthong D, Miranda AM, Udomsantisuk N, Ellis K, van Leeuwen J, *et al*. A more reliable PCR for detection of *Mycobacterium tuberculosis* in clinical samples. *J Clin Microbiol* 1994;**32**:672–8.

351. Li JY, Lo ST, Ng CS. Molecular detection of *Mycobacterium tuberculosis* in tissues showing granulomatous inflammation without demonstrable acid-fast bacilli. *Diagn Mol Pathol* 2000;**9**:67–74.

352. Montenegro SH, Gilman RH, Sheen P, Cama R, Caviedes L, Hopper T, *et al*. Improved detection of *Mycobacterium tuberculosis* in Peruvian children by use of a heminested IS6110 polymerase chain reaction assay. *Clin Infect Dis* 2003;**36**:16–23.

353. Nastasi A, Mammina C, Lucia DC. Contribution of nucleic acid amplification to diagnosis of pulmonary tuberculosis. *Ann Ig* 1997;**9**:347–52.

354. Pao CC, Yen TSB, You J-B, Maa J-S, Fiss EH, Chang C-H. Detection and identification of *Mycobacterium tuberculosis* by DNA amplification. *J Clin Microbiol* 1990;**28**:1877–80.

355. Schluger NW, Kinney D, Harkin TJ, Rom WN. Clinical utility of the polymerase chain reaction in the diagnosis of infections due to *Mycobacterium tuberculosis*. *Chest* 1994;**105**:1116–21.

356. Shim JJ, Cheong HJ, Kang EY, In KH, Yoo SH, Kang KH. Nested polymerase chain reaction for detection of *Mycobacterium tuberculosis* in solitary pulmonary nodules. *Chest* 1998;**113**:20–4.

357. Tan MF, Ng WC, Chan SH, Tan WC. Comparative usefulness of PCR in the detection of *M. tuberculosis* in different clinical specimens. *J Med Microbiol* 1997;**46**:164–9.

358. Tan YK, Lee AS, Khoo KL, Ong SY, Wong SY, Ong YY. Rapid mycobacterial tuberculosis detection in bronchoalveolar lavage samples by polymerase chain reaction in patients with upper lobe infiltrates and bronchiectasis. *Ann Acad Med Singapore* 1999;**28**:205–8.

359. Tansuphasiri U. Comparison of microplate hybridization with gel electrophoresis and dot blot hybridization for the rapid detection of *Mycobacterium tuberculosis* PCR products. *Southeast Asian J Trop Med Public Health* 2002;**33**:136–46.

360. Wong CF, Yew WW, Chan CY, Au LY, Cheung SW, Cheng AF. Rapid diagnosis of smear-negative pulmonary tuberculosis via fibreoptic bronchoscopy: utility of polymerase chain reaction in bronchial aspirates as an adjunct to transbronchial biopsies. *Respir Med* 1998;**92**:815–19.

361. Zambardi G, Roure C, Boujaafar N, Fouque B, Freney J, Fleurette J. Comparison of three primer sets for the detection of *Mycobacterium tuberculosis* in clinical samples by polymerase chain reaction. *Ann Biol Clin (Paris)* 1993;**51**:893–7.

362. Bergmann JS, Keating WE, Woods GL. Clinical evaluation of the BDProbeTec ET system for rapid detection of *Mycobacterium tuberculosis*. *J Clin Microbiol* 2000;**38**:863–5.

363. Huang TS, Huang WK, Lee SS, Tu HZ, Chang SH, Liu YC. Rapid detection of pulmonary tuberculosis using the BDProbeTEC ET Mycobacterium Tuberculosis Complex Direct Detection Assay (DTB). *Diagn Microbiol Infect Dis* 2003;**46**:29–33.

364. Alcaide F, Gali N, Dominguez J, Berlanga P, Blanco S, Orus P, *et al*. Usefulness of a new mycobacteriophage-based technique for rapid diagnosis of pulmonary tuberculosis. *J Clin Microbiol* 2003;**41**:2867–71.

365. Chander J, Subrahmanyan S, Gupta R. Sensitivity of EIA in the diagnosis of tuberculosis using 38-kDa antigen. *J Indian Med Assoc* 1996;**94**:376–8.

366. Charpin D, Herbault H, Gevaudan MJ, Saadjian M, De Micco P, Arnaud A, *et al*. Value of ELISA using A60 antigen in the diagnosis of active pulmonary tuberculosis. *Am Rev Respir Dis* 1990; **142**:380–4.

367. Luh KT, Yu CJ, Yang PC, Lee LN. Tuberculosis antigen A60 serodiagnosis in tuberculous infection: application in extrapulmonary and smear-negative pulmonary tuberculosis. *Respirology* 1996;**1**:145–51.

368. McConkey SJ, Youssef FG, Azem E, Frenck RW, Weil GJ. Evaluation of a rapid-format antibody test and the tuberculin skin test for diagnosis of tuberculosis in two contrasting endemic settings. *Int J Tuberc Lung Dis* 2002;**6**:246–52.

369. Somi GR, O'Brien RJ, Mfinanga GS, Ipuge YA. Evaluation of the MycoDot test in patients with suspected tuberculosis in a field setting in Tanzania. *Int J Tuberc Lung Dis* 1999;**3**:231–8.

370. Chan ED, Reves R, Belisle JT, Brennan PJ, Hahn WE. Diagnosis of tuberculosis by a visually detectable immunoassay for lipoarabinomannan. *Am J Respir Crit Care Med* 2000;**161**:1713–19.

371. Levy H, Wadee AA, Feldman C, Rabson AR. Enzyme-linked immunosorbent assay for the detection of antibodies against *Mycobacterium tuberculosis* in bronchial washings and serum. *Chest* 1988;**93**:762–6.

372. Morris CD, Nell H. Epidemic of pulmonary tuberculosis in geriatric homes. *S Afr Med J* 1988; **74**:117–20.

373. Nicholls AC, Horsfield K. Serological diagnosis of tuberculosis: a report of 12 months' clinical experience. *Thorax* 1976;**31**:289–93.

374. Silva VMC, Kanaujia G, Gennaro ML, Menzies D. Factors associated with humoral response to ESAT-6, 38 kDa and 14 kDa in patients with a spectrum of tuberculosis. *Int J Tuberc Lung Dis* 2003; **7**:478–84.

375. Tessema TA, Bjune G, Hamasur B, Svenson S, Syre H, Bjorvatn B. Circulating antibodies to lipoarabinomannan in relation to sputum microscopy, clinical features and urinary anti-lipoarabinomannan detection in pulmonary tuberculosis. *Scand J Infect Dis* 2002;**34**:97–103.

376. Zeiss CR, Kalish SB, Erlich KS, Levitz D, Metzger E, Radin R, *et al*. IgG antibody to purified protein derivative by enzyme-linked immunosorbent assay in the diagnosis of pulmonary tuberculosis. *Am Rev Respir Dis* 1984; **130**:845–8.

377. Bemer-Melchior P, Germaud P, Drugeon HB. Diagnosis of extrapulmonary tuberculosis by a commercial polymerase chain reaction kit. *Pathol Biol (Paris)* 1998;**46**:597–603.

378. Gamboa F, Manterola JM, Vinado B, Matas L, Gimenez M, Lonca J, *et al*. Direct detection of *Mycobacterium tuberculosis* complex in nonrespiratory specimens by Gen-Probe Amplified Mycobacterium Tuberculosis Direct Test. *J Clin Microbiol* 1997;**35**:307–10.

379. Gamboa F, Fernandez G, Padilla E, Manterola JM, Lonca J, Cardona PJ, *et al*. Comparative evaluation of initial and new versions of the Gen-Probe Amplified Mycobacterium Tuberculosis Direct Test for direct detection of *Mycobacterium tuberculosis* in respiratory and nonrespiratory specimens. *J Clin Microbiol* 1998;**36**:684–9.

380. Gamboa F, Dominguez J, Padilla E, Manterola JM, Gazapo E, Lonca J, *et al*. Rapid diagnosis of extrapulmonary tuberculosis by ligase chain reaction amplification. *J Clin Microbiol* 1998; **36**:1324–9.

381. Honore-Bouakline S, Vincensini JP, Giacuzzo V, Lagrange PH, Herrmann JL. Rapid diagnosis of extrapulmonary tuberculosis by PCR: impact of sample preparation and DNA extraction. *J Clin Microbiol* 2003;**41**:2323–9.

382. Shah S, Miller A, Mastellone A, Kim K, Colaninno P, Hochstein L, *et al*. Rapid diagnosis of tuberculosis in various biopsy and body fluid specimens by the AMPLICOR *Mycobacterium tuberculosis* polymerase chain reaction test. *Chest* 1998;**113**:1190–4.

383. Brisson-Noel A, Aznar C, Chureau C, Nguyen S, Pierre C, Bartoli M, *et al*. Diagnosis of tuberculosis by DNA amplification in clinical practice evaluation. *Lancet* 1991;**338**:364–6.

384. Hardman WJ, Benian GM, Howard T, McGowan JEJ, Metchock B, Murtagh JJ. Rapid detection of mycobacteria in inflammatory necrotizing granulomas from formalin-fixed, paraffin-embedded tissue by PCR in clinically high-risk patients with acid-fast stain and culture-negative tissue biopsies. *Am J Clin Pathol* 1996; **106**:384–9.

385. Kaltwasser G, Garcia S, Salinas AM, Montiel F. Enzymatic DNA amplification (PCR) in the diagnosis of extrapulmonary *Mycobacterium tuberculosis* infection. *Mol Cell Probes* 1993;**7**:465–70.

386. Kolk AH, Kox LF, van Leeuwen J, Kuijper S, Jansen HM. Clinical utility of the polymerase chain reaction in the diagnosis of extrapulmonary tuberculosis. *Eur Respir J* 1998;**11**:1222–6.

387. Portillo-Gomez L, Morris SL, Panduro A. Rapid and efficient detection of extra-pulmonary *Mycobacterium tuberculosis* by PCR analysis. *Int J Tuberc Lung Dis* 2000;**4**:361–70.

388. Salian NV, Rish JA, Eisenach KD, Cave MD, Bates JH. Polymerase chain reaction to detect *Mycobacterium tuberculosis* in histologic specimens. *Am J Respir Crit Care Med* 1998;**158**:1150–5.

389. Goode D. A serological test for tuberculosis. *J R Army Med Corps* 1980;**126**:33–6.

390. Ivanyi J, Krambovitis E, Keen M. Evaluation of a monoclonal antibody (TB72) based serological test for tuberculosis. *Clin Exp Immunol* 1983;**54**:337–45.

391. de Wit D, Maartens G, Steyn L. A comparative study of the polymerase chain reaction and conventional procedures for the diagnosis of tuberculous pleural effusion. *Tuber Lung Dis* 1992; **73**:262–7.

392. Parandaman V, Narayanan S, Narayanan PR. Utility of polymerase chain reaction using two probes for rapid diagnosis of tubercular pleuritis in comparison to conventional methods. *Indian J Med Res* 2000;**112**:47–51.

393. Reechaipichitkul W, Lulitanond V, Sungkeeree S, Patjanasoontorn B. Rapid diagnosis of tuberculous pleural effusion using polymerase chain reaction. *Southeast Asian J Trop Med Public Health* 2000; **31**:509–14.

394. Takagi N, Hasegawa Y, Ichiyama S, Shibagaki T, Shimokata K. Polymerase chain reaction of pleural biopsy specimens for rapid diagnosis of tuberculous pleuritis. *Int J Tuberc Lung Dis* 1998; **2**:338–41.

395. Villena V, Rebollo MJ, Aguado JM, Galan A, Lopez-Encuentra A, Palenque E. Polymerase chain reaction for the diagnosis of pleural tuberculosis in immunocompromised and immunocompetent patients. *Clin Infect Dis* 1998;**26**:212–14.

396. Aoki Y, Katoh O, Nakanishi Y, Kuroki S, Yamada H. A comparison study of IFN-gamma, ADA, and CA125 as the diagnostic parameters in tuberculous pleuritis. *Respir Med* 1994;**88**:139–43.

397. Banales JL, Pineda PR, Fitzgerald JM, Rubio H, Selman M, Salazar-Lezama M. Adenosine deaminase in the diagnosis of tuberculous pleural effusions. A report of 218 patients and review of the literature. *Chest* 1991;**99**:355–7.

398. Blake J, Berman P. The use of adenosine deaminase assays in the diagnosis of tuberculosis. *S Afr Med J* 1982;**62**:19–21.

399. Burgess LJ, Maritz FJ, Le Roux I, Taljaard JJ. Combined use of pleural adenosine deaminase with lymphocyte/neutrophil ratio. Increased specificity for the diagnosis of tuberculous pleuritis. *Chest* 1996;**109**:414–19.

400. Caballero M, Ruiz R, Marquez DP, Seco M, Borque L, Escanero JF. Development of a microparticle-enhanced nephelometric immunoassay for quantitation of human lysozyme in pleural effusion and plasma. *J Clin Lab Anal* 1999;**13**:301–7.

401. Fontan-Bueso J, Verea-Hernando H, Garcia-Buela JP, Dominguez-Juncal L, Martin-Egana MT, Montero-Martinez MC. Diagnostic value of simultaneous determination of pleural adenosine deaminase and pleural lysozyme/serum lysozyme ratio in pleural effusions. *Chest* 1988; **93**:303–7.

402. Goulart de Oliveira H, Rossatto ER, Prolla JC. Pleural fluid adenosine deaminase and lymphocyte proportion: Clinical usefulness in the diagnosis of tuberculosis. *Cytopathology* 1994; **5**:27–32.

403. Gupta DK, Suri JC, Goel A. Efficacy of adenosine deaminase in the diagnosis of pleural effusions. *Indian J Chest Dis Allied Sci* 1990;**32**:205–8.

404. Hsu WH, Chiang CD, Huang PL. Diagnostic value of pleural adenosine deaminase in tuberculous effusions of immunocompromised hosts. *J Formos Med Assoc* 1993;**92**:668–70.

405. Kuralay F, Comlekci A. Adenosine deaminase activity: to be an useful marker in distinguishing pleural effusions due to malignancy from tuberculosis. *Biochem Soc Trans* 1998;**26**:S163.

406. Maritz FJ, Malan C, Le Roux I. Adenosine deaminase estimations in the differentiation of pleural effusions. *S Afr Med J* 1982;**62**:556–8.

407. Momi H, Matsuyama W, Inoue K, Kawabata M, Arimura K, Fukunaga H, *et al*. Vascular endothelial growth factor and proinflammatory cytokines in pleural effusions. *Respir Med* 2002;**96**:817–22.

408. Ocana I, Martinez-Vazquez JM, Ribera E, Segura RM, Pascual C. Adenosine deaminase activity in the diagnosis of lymphocytic pleural effusions of tuberculous, neoplastic and lymphomatous origin. *Tubercle* 1986;**67**:141–5.

409. Orphanidou D, Gaga M, Rasidokis A, Dimakou K, Toumbis M, Latsi P. Tumor necrosis factor, interleukin 1 and adenosine deaminase in tuberculous pleural effusion. *Respir Med* 1996; **90**:95–8.

410. Perez-Rodriguez E, Perez-Walton IJ, Sanchez-Hernandez JJ, Pallares E, Rubi J, Jimenez-Castro D *et al*. ADA1/ADAp ratio in pleural tuberculosis: an excellent diagnostic parameter in pleural fluid. *Respir Med* 1999;**93**:816–21.

411. Pettersson T. Adenosine deaminase EC-3.5.4.4 in the diagnosis of pleural effusions. *Acta Med Scand* 1984;**215**:299–304.

412. Reechaipichitkul W, Kawamatawong T, Teerajetgul Y, Patjanasoontorn B. Diagnostic role of pleural fluid adenosine deaminase in tuberculous pleural effusion. *Southeast Asian J Trop Med Public Health* 2001;**32**:383–9.

413. Richter C, Perenboom R, Swai AB, Kitinya J, Mtoni I, Chande H, *et al*. Diagnosis of tuberculosis in patients with pleural effusion in an area of HIV infection and limited diagnostic facilities. *Trop Geogr Med* 1994;**46**:293–7.

414. San Jose E, Valdes L, Sarandeses A, Alvarez D, Chomon B. Diagnostic value of adenosine deaminase and lysozyme in tuberculous pleurisy. *Clin Chim Acta* 1992;**209**:73–81.

415. Sharma SK, Suresh V, Mohan A, Kaur P, Saha P, Kumar A, *et al*. A prospective study of sensitivity and specificity of adenosine deaminase estimation in the diagnosis of tuberculosis pleural effusion. *Indian J Chest Dis Allied Sci* 2001;**43**:149–55.

416. Sinha PK, Sinha BB, Sinha AR. Adenosine deaminase activity as a diagnostic index of pleural effusion. *J Indian Med Assoc* 1987;**85**:11–13.

417. Strankinga WF, Nauta JJ, Straub JP, Stam J. Adenosine deaminase activity in tuberculous pleural effusions: a diagnostic test. *Tubercle* 1987; **68**:137–40.

418. Valdes L, San Jose E, Alvarez D, Sarandeses A, Pose A, Chomon B, *et al*. Diagnosis of tuberculous pleurisy using the biologic parameters adenosine deaminase, lysozyme, and interferon gamma. *Chest* 1993;**103**:458–65.

419. Valdes L, Alvarez D, San Jose E, Juanatey JR, Pose A, Valle JM, *et al*. Value of adenosine deaminase in the diagnosis of tuberculous pleural effusions in young patients in a region of high prevalence of tuberculosis. *Thorax* 1995;**50**:600–3.

420. Valdes L, San Jose E, Alvarez D, Valle JM. Adenosine deaminase (ADA) isoenzyme analysis in pleural effusions: diagnostic role, and relevance to the origin of increased ADA in tuberculous pleurisy. *Eur Respir J* 1996;**9**:747–51.

421. Villena V, Navarro-Gonzalvez JA, Garcia-Benayas C, Manzanos JA, Echave J, Lopez-Encuentra A, *et al*. Rapid automated determination of adenosine deaminase and lysozyme for differentiating tuberculous and nontuberculous pleural effusions. *Clin Chem* 1996;**42**:218–21.

422. Ribera E, Ocana I, Martinez-Vasquez JM, Rossell M, Espanol T, Ruibal A. High level of interferon gamma in tuberculous pleural effusion. *Chest* 1988;**93**:308–11.

423. Silva-Mejias C, Gamboa-Antinolo F, Lopez-Cortes LF, Cruz-Ruiz M, Pachon J. Interleukin-1 beta in pleural fluids of different etiologies. Its role as inflammatory mediator in empyema. *Chest* 1995;**108**:942–5.

424. Wongtim S, Silachamroon U, Ruxringtham K, Udompanich V, Limthongkul S, Charoenlap P, *et al*. Interferon gamma for diagnosing tuberculous pleural effusions. *Thorax* 1999;**54**:921–4.

425. Asseo P, Tracopoulos GD, Kotsovoulous-Fouskaki V. Lysozme (muramidase) in pleural effusions and serum. *Am J Clin Pathol* 1982;**78**:763–7.

426. Klockars M, Pettersson T, Riska H, Hellstrom PE, Norhagen A. Pleural fluid lysozome in human disease. *Arch Intern Med* 1979;**139**:73–7.

427. Yorgancioglu A, Akin M, Dereli S, Aktogu S, Ilis Z, Sezgin A. The diagnostic value of tuberculostearic acid in tuberculous pleural effusions. *Monaldi Arch Chest Dis* 1996;**51**:108–11.

428. Bonington A, Strang JI, Klapper PE, Hood SV, Rubombora W, Penny M, *et al*. Use of Roche AMPLICOR *Mycobacterium tuberculosis* PCR in early diagnosis of tuberculous meningitis. *J Clin Microbiol* 1998;**36**:1251–4.

429. Bonington A, Strang JI, Klapper PE, Hood SV, Parish A, Swift PJ, *et al*. TB PCR in the early diagnosis of tuberculous meningitis: evaluation of the Roche semi-automated COBAS Amplicor MTB test with reference to the manual Amplicor MTB PCR test. *Tuberc Lung Dis* 2000;**80**:191–6.

430. Brienze VM, Tonon AP, Pereira FJ, Liso E, Tognola WA, dos-Santos MA, *et al*. Low sensitivity of polymerase chain reaction for diagnosis of tuberculous meningitis in southeastern Brazil. *Rev Soc Bras Med Trop* 2001;**34**:389–93.

431. Chedore P, Jamieson FB. Rapid molecular diagnosis of tuberculous meningitis using the Gen-probe Amplified Mycobacterium Tuberculosis direct test in a large Canadian public health laboratory. *Int J Tuberc Lung Dis* 2002;**6**:913–19.

432. Lang AM, Feris-Iglesias J, Pena C, Sanchez JF, Stockman L, Rys P, *et al*. Clinical evaluation of the Gen-Probe Amplified Direct Test for detection of *Mycobacterium tuberculosis complex* organisms in cerebrospinal fluid. *J Clin Microbiol* 1998; **36**:2191–4.

433. Rajo MC, Perez-del-Molino ML, Lado-Lado FL, Lopez MJ, Prieto E, Pardo F. Rapid diagnosis of tuberculous meningitis by ligase chain reaction amplification. *Scand J Infect Dis* 2002;**34**:14–16.

434. Ahuja GK, Mohan KK, Prasad K, Behari M. Diagnostic criteria for tuberculous meningitis and their validation. *Tuberc Lung Dis* 1994;**75**:149–52.

435. Caws M, Wilson SM, Clough C, Drobniewski F. Role of IS6110-targeted PCR, culture, biochemical, clinical, and immunological criteria for diagnosis of tuberculous meningitis. *J Clin Microbiol* 2000;**38**:3150–5.

436. Hooker JA, Muhindi DW, Amayo EO, Mc'ligeyo SO, Bhatt KM, Odhiambo JA. Diagnostic utility of cerebrospinal fluid studies in patients with clinically suspected tuberculous meningitis. *Int J Tuberc Lung Dis* 2003;**7**:787–96.

437. Kox LF, Kuijper S, Kolk AH. Early diagnosis of tuberculous meningitis by polymerase chain reaction. *Neurology* 1995;**45**:2228–32.

438. Narayanan S, Parandaman V, Narayanan PR, Venkatesan P, Girish C, Mahadevan S, *et al*. Evaluation of PCR using TRC(4) and IS6110 primers in detection of tuberculous meningitis. *J Clin Microbiol* 2001;**39**:2006–8.

439. Nguyen LN, Kox LF, Pham LD, Kuijper S, Kolk AH. The potential contribution of the polymerase chain reaction to the diagnosis of tuberculous meningitis. *Arch Neurol* 1996;**53**:771–6.

440. Gambhir IS, Mehta M, Singh DS, Khanna HD. Evaluation of CSF-adenosine deaminase activity in tubercular meningitis. *J Assoc Physicians India* 1999;**47**:192–4.

441. Mann MD, Macfarlane CM, Verburg CJ, Wiggelinkhuizen J. The bromide partition test and CSF adenosine deaminase activity in the diagnosis of tuberculosis meningitis in children. *S Afr Med J* 1982;**62**:431–3.

442. Pettersson T, Klockars M, Weber TH, Somer H. Diagnostic value of cerebrospinal fluid adenosine deaminase determination. *Scand J Infect Dis* 1991;**23**:97–100.

443. Rohani MY, Cheong YM, Rani JM. The use of adenosine deaminase activity as a biochemical marker for the diagnosis of tuberculous meningitis. *Malays J Pathol* 1995;**17**:67–71.

444. Bal V, Kamat RS, Kamat J, Kandoth P. Enzyme-linked immunosorbent assay for mycobacterial antigens. *Indian J Med Res* 1983;**78**:477–83.

445. Donald PR, Cooper RC. Enzyme-linked immunosorbent assay for the detection of mycobacterial antigens in the cerebrospinal fluid in tuberculous meningitis. *S Afr Med J* 1987; **71**:699–700.

446. Rimek D, Tyagi S, Kappe R. Performance of an IS6110-based PCR assay and the COBAS AMPLICOR MTB PCR system for detection of *Mycobacterium tuberculosis complex* DNA in human lymph node samples. *J Clin Microbiol* 2002; **40**:3089–92.

447. Manitchotpisit B, Kunachak S, Kulapraditharom B, Sura T. Combined use of fine needle aspiration cytology and polymerase chain reaction in the diagnosis of cervical tuberculous lymphadenitis. *J Med Assoc Thai* 1999;**82**:363–8.

448. Narayanan S, Parandaman V, Rehman F, Srinivasan C, Gomathy D, Kumaraswami V, *et al*. Comparative evaluation of PCR using IS6110 and a new target in the detection of tuberculous lymphadenitis. *Curr Sci (Bangalore)* 2000;**78**:1367–70.

449. Bhargava DK, Gupta M, Nijhawan S, Dasarathy S, Kushwaha AK. Adenosine deaminase (ADA) in peritoneal tuberculosis: diagnostic value in ascitic fluid and serum. *Tubercle* 1990;**71**:121–6.

450. Burgess LJ, Swanepoel CG, Taljaard JJ. The use of adenosine deaminase as a diagnostic tool for peritoneal tuberculosis. *Tuberculosis (Edinb)* 2001; **81**:243–8.

451. Dwivedi M, Misra SP, Misra V, Kumar R. Value of adenosine deaminase estimation in the diagnosis of tuberculous ascites. *Am J Gastroenterol* 1990; **85**:1123–5.

452. Koh KK, In HH, Lee KH, Kim EJ, Cho CH, Cho SK, *et al*. New scoring system using tumor markers in diagnosing patients with moderate pericardial effusions. *Int J Cardiol* 1997;**61**:5–13.

453. Komsuoglu B, Goldeli O, Kulan K, Komsuoglu SS. The diagnostic and prognostic value of adenosine deaminase in tuberculous pericarditis. *Eur Heart J* 1995;**16**:1126–30.

454. van Vollenhoven P, Heyns CF, de Beer PM, Whitaker P, van Helden PD, Victor T. Polymerase chain reaction in the diagnosis of urinary tract tuberculosis. *Urol Res* 1996;**24**:107–11.

455. Kanchana MV, Cheke D, Natyshak I, Connor B, Warner A, Martin T. Evaluation of the BACTEC MGIT 960 system for the recovery of mycobacteria. *Diagn Microbiol Infect Dis* 2000; **37**:31–6.

456. Lu DS, Heeren B, Dunne WM. Comparison of the automated mycobacteria growth indicator tube system (BACTEC 960/MGIT) with Lowenstein–Jensen medium for recovery of mycobacteria from clinical specimens. *Am J Clin Pathol* 2002;**118**:542–5.

457. Benjamin WHJ, Waites KB, Beverly A, Gibbs L, Waller M, Nix S, *et al*. Comparison of the MB/BacT system with a revised antibiotic supplement kit to the BACTEC 460 system for detection of mycobacteria in clinical specimens. *J Clin Microbiol* 1998;**36**:3234–8.

458. Brunello F, Favari F, Fontana R. Comparison of the MB/BacT and BACTEC 460 TB systems for recovery of mycobacteria from various clinical specimens. *J Clin Microbiol* 1999;**37**:1206–9.

459. Gonzalez N, Torres MJ, Aznar J, Palomares JC. Molecular analysis of rifampin and isoniazid resistance of *Mycobacterium tuberculosis* clinical isolates in Seville, Spain. *Tubercle Lung Dis* 1999; **79**:187–90.

460. Harris G, Rayner A, Blair J, Watt B. Comparison of three isolation systems for the culture of mycobacteria from respiratory and non-respiratory samples. *J Clin Pathol* 2000;**53**:615–18.

461. Palacios JJ, Ferro J, Ruiz-Palma N, Garcia JM, Villar H, Rodriguez J, *et al*. Fully automated liquid culture system compared with Lowenstein–Jensen solid medium for rapid recovery of mycobacteria from clinical samples. *Eur J Clin Microbiol Infect Dis* 1999;**18**:265–73.

462. Piersimoni C, Scarparo C, Callegaro A, Tosi CP, Nista D, Bornigia S, *et al*. Comparison of MB/Bact alert 3D system with radiometric BACTEC system and Lowenstein–Jensen medium for recovery and identification of mycobacteria from clinical specimens: a multicenter study. *J Clin Microbiol* 2001;**39**:651–7.

463. Scarparo C, Piccoli P, Rigon A, Ruggiero G, Ricordi P, Piersimoni C. Evaluation of the BACTEC MGIT 960 in comparison with BACTEC 460 TB for detection and recovery of mycobacteria from clinical specimens. *Diagn Microbiol Infect Dis* 2002;**44**:157–61.

464. Roggenkamp A, Hornef MW, Masch A, Aigner B, Autenrieth IB, Heesemann J. Comparison of MB/BacT and BACTEC 460

TB systems for recovery of mycobacteria in a routine diagnostic laboratory. *J Clin Microbiol* 1999;**37**:3711–12.

465. Rohner P, Ninet B, Metral C, Emler S, Auckenthaler R. Evaluation of the MB/BacT system and comparison to the BACTEC 460 system and solid media for isolation of mycobacteria from clinical specimens. *J Clin Microbiol* 1997;**35**:3127–31.

466. Somoskovi A, Kodmon C, Lantos A, Bartfai Z, Tamasi L, Fuzy J, *et al*. Comparison of recoveries of *Mycobacterium tuberculosis* using the automated BACTEC MGIT 960 system, the BACTEC 460 TB system, and Lowenstein–Jensen medium. *J Clin Microbiol* 2000;**38**:2395–7.

467. Van Griethuysen AJ, Jansz A, Buiting AG. Comparison of fluorescent BACTEC 9000 MB system, Septi-Chek AFB system, and Lowenstein–Jensen medium for detection of mycobacteria. *J Clin Microbiol* 1996;**34**:2391–4.

468. Tortoli E, Cichero P, Chirillo MG, Gismondo MR, Bono L, Gesu G, *et al*. Multicenter comparison of ESP Culture System II with BACTEC 460TB and with Lowenstein–Jensen medium for recovery of mycobacteria from different clinical specimens, including blood. *J Clin Microbiol* 1998;**36**:1378–81.

469. Williams-Bouyer N, Yorke R, Lee HI, Woods GL. Comparison of the BACTEC MGIT 960 and ESP culture system II for growth and detection of mycobacteria. *J Clin Microbiol* 2000;**38**:4167–70.

470. Woods GL, Fish G, Plaunt M, Murphy T. Clinical evaluation of Difco ESP culture system II for growth and detection of mycobacteria. *J Clin Microbiol* 1997;**35**:121–4.

471. *British National Formulary*. London: BMJ Books; 2003.

This version of HTA monograph volume 11, number 3 does not include the 122 pages of appendices. This is to save download time from the HTA website.

The printed version of this monograph also excludes the appendices.

View/download the appendices (617 kbytes).

Health Technology Assessment Programme

Prioritisation Strategy Group

Members

HTA Commissioning Board

Members

Current and past membership details of all HTA 'committees' are available from the HTA website (www.hta.ac.uk)

Diagnostic Technologies & Screening Panel

Members

Chair,
Dr Ron Zimmern, Director of the Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

Ms Norma Armston, Freelance Consumer Advocate, Bolton

Professor Max Bachmann, Professor of Health Care Interfaces, Department of Health Policy and Practice, University of East Anglia

Professor Rudy Bilous Professor of Clinical Medicine & Consultant Physician, The Academic Centre, South Tees Hospitals NHS Trust

Dr Paul Cockcroft, Consultant Medical Microbiologist and Clinical Director of Pathology, Department of Clinical Microbiology, St Mary's Hospital, Portsmouth

Professor Adrian K Dixon, Professor of Radiology, University Department of Radiology, University of Cambridge Clinical School

Dr David Elliman, Consultant Paediatrician/ Hon. Senior Lecturer, Population Health Unit, Great Ormond St. Hospital, London

Professor Glyn Elwyn, Primary Medical Care Research Group, Swansea Clinical School, University of Wales Swansea

Professor Paul Glasziou, Director, Centre for Evidence-Based Practice, University of Oxford

Dr Jennifer J Kurinczuk, Consultant Clinical Epidemiologist, National Perinatal Epidemiology Unit, Oxford

Dr Susanne M Ludgate, Medical Director, Medicines & Healthcare Products Regulatory Agency, London

Mr Stephen Pilling, Director, Centre for Outcomes, Research & Effectiveness, Joint Director, National Collaborating Centre for Mental Health, University College London

Dr Phil Shackley, Senior Lecturer in Health Economics, School of Population and Health Sciences, University of Newcastle upon Tyne

Dr Margaret Somerville, Director of Public Health Learning, Peninsula Medical School, University of Plymouth

Dr Graham Taylor, Scientific Director & Senior Lecturer, Regional DNA Laboratory, The Leeds Teaching Hospitals

Professor Lindsay Wilson Turnbull, Scientific Director, Centre for MR Investigations & YCR Professor of Radiology, University of Hull

Professor Martin J Whittle, Clinical Co-director, National Co-ordinating Centre for Women's and Childhealth

Dr Dennis Wright, Consultant Biochemist & Clinical Director, Pathology & The Kennedy Galton Centre, Northwick Park & St Mark's Hospitals, Harrow

Pharmaceuticals Panel

Members

Chair,
Dr John Reynolds, Chair Division A, The John Radcliffe Hospital, Oxford Radcliffe Hospitals NHS Trust

Ms Anne Baileff, Consultant Nurse in First Contact Care, Southampton City Primary Care Trust, University of Southampton

Professor Imti Choonara, Professor in Child Health, Academic Division of Child Health, University of Nottingham

Dr Robin Ferner, Consultant Physician and Director, West Midlands Centre for Adverse Drug Reactions, City Hospital NHS Trust, Birmingham

Professor John Geddes, Professor of Epidemiological Psychiatry, University of Oxford

Mrs Barbara Greggains, Non-Executive Director, Greggains Management Ltd

Dr Bill Gutteridge, Medical Adviser, National Specialist Commissioning Advisory Group (NSCAG), London

Mrs Sharon Hart, Consultant Pharmaceutical Adviser, Reading

Dr Jon Karnon, Senior Research Fellow, Health Economics and Decision Science, University of Sheffield

Dr Yoon Loke, Senior Lecturer in Clinical Pharmacology, University of East Anglia

Ms Barbara Meredith, Lay Member, Epsom

Dr Andrew Prentice, Senior Lecturer and Consultant Obstetrician & Gynaecologist, Department of Obstetrics & Gynaecology, University of Cambridge

Dr Frances Rotblat, CPMP Delegate, Medicines & Healthcare Products Regulatory Agency, London

Dr Martin Shelly, General Practitioner, Leeds

Mrs Katrina Simister, Assistant Director New Medicines, National Prescribing Centre, Liverpool

Dr Richard Tiner, Medical Director, Medical Department, Association of the British Pharmaceutical Industry, London

Therapeutic Procedures Panel

Members

Chair,
Professor Bruce Campbell, Consultant Vascular and General Surgeon, Department of Surgery, Royal Devon & Exeter Hospital

Dr Mahmood Adil, Deputy Regional Director of Public Health, Department of Health, Manchester

Dr Aileen Clarke, Consultant in Public Health, Public Health Resource Unit, Oxford

Professor Matthew Cooke, Professor of Emergency Medicine, Warwick Emergency Care and Rehabilitation, University of Warwick

Ms Amelia Curwen, Executive Director of Policy, Services and Research, Asthma UK, London

Mr Mark Emberton, Senior Lecturer in Oncological Urology, Institute of Urology, University College Hospital

Professor Gene Feder, Professor of Primary Care R&D, Department of General Practice and Primary Care, Barts & the London, Queen Mary's School of Medicine and Dentistry, London

Professor Paul Gregg, Professor of Orthopaedic Surgical Science, Department of General Practice and Primary Care, South Tees Hospital NHS Trust, Middlesbrough

Ms Bec Hanley, Co-Director, TwoCan Associates, Hurstpierpoint

Ms Maryann L Hardy, Lecturer, Division of Radiography, University of Bradford

Professor Alan Horwich, Director of Clinical R&D, Academic Department of Radiology, The Institute of Cancer Research, London

Dr Simon de Lusignan, Senior Lecturer, Primary Care Informatics, Department of Community Health Sciences, St George's Hospital Medical School, London

Dr Peter Martin, Consultant Neurologist, Addenbrooke's Hospital, Cambridge

Professor Neil McIntosh, Edward Clark Professor of Child Life & Health, Department of Child Life & Health, University of Edinburgh

Professor James Neilson, Professor of Obstetrics and Gynaecology, Department of Obstetrics and Gynaecology, University of Liverpool

Dr John C Pounsford, Consultant Physician, Directorate of Medical Services, North Bristol NHS Trust

Dr Karen Roberts, Nurse Consultant, Queen Elizabeth Hospital, Gateshead

Dr Vimal Sharma, Consultant Psychiatrist/Hon. Senior Lecturer, Mental Health Resource Centre, Cheshire and Wirral Partnership NHS Trust, Wallasey

Professor Norman Waugh, Professor of Public Health, Department of Public Health, University of Aberdeen

Professor Scott Weich, Professor of Psychiatry, Division of Health in the Community, University of Warwick

Disease Prevention Panel

Members

Chair,
Dr Edmund Jessop, Medical Adviser, National Specialist Commissioning Advisory Group (NSCAG), London

Mrs Sheila Clark, Chief Executive, St James's Hospital, Portsmouth

Mr Richard Copeland, Lead Pharmacist: Clinical Economy/Interface, Wansbeck General Hospital, Northumberland

Dr Elizabeth Fellow-Smith, Medical Director, West London Mental Health Trust, Middlesex

Mr Ian Flack, Director PPI Forum Support, Council of Ethnic Minority Voluntary Sector Organisations, Stratford

Dr John Jackson, General Practitioner, Newcastle upon Tyne

Mrs Veronica James, Chief Officer, Horsham District Age Concern, Horsham

Professor Mike Kelly, Director, Centre for Public Health Excellence, National Institute for Health and Clinical Excellence, London

Professor Yi Mien Koh, Director of Public Health and Medical Director, London NHS (North West London Strategic Health Authority), London

Ms Jeanett Martin, Director of Clinical Leadership & Quality, Lewisham PCT, London

Dr Chris McCall, General Practitioner, Dorset

Dr David Pencheon, Director, Eastern Region Public Health Observatory, Cambridge

Dr Ken Stein, Senior Clinical Lecturer in Public Health, Director, Peninsula Technology Assessment Group, University of Exeter, Exeter

Dr Carol Tannahill, Director, Glasgow Centre for Population Health, Glasgow

Professor Margaret Thorogood, Professor of Epidemiology, University of Warwick, Coventry

Dr Ewan Wilkinson, Consultant in Public Health, Royal Liverpool University Hospital, Liverpool

Expert Advisory Network

Members

Professor Douglas Altman, Director of CSM & Cancer Research UK Med Stat Gp, Centre for Statistics in Medicine, University of Oxford, Institute of Health Sciences, Headington, Oxford

Professor John Bond, Director, Centre for Health Services Research, University of Newcastle upon Tyne, School of Population & Health Sciences, Newcastle upon Tyne

Professor Andrew Bradbury, Professor of Vascular Surgery, Birmingham Heartlands Hospital, Birmingham

Mr Shaun Brogan, Chief Executive, Ridgeway Primary Care Group, Aylesbury

Mrs Stella Burnside OBE, Chief Executive, Office of the Chief Executive. Trust Headquarters, Altnagelvin Hospitals Health & Social Services Trust, Altnagelvin Area Hospital, Londonderry

Ms Tracy Bury, Project Manager, World Confederation for Physical Therapy, London

Professor Iain T Cameron, Professor of Obstetrics and Gynaecology and Head of the School of Medicine, University of Southampton

Dr Christine Clark, Medical Writer & Consultant Pharmacist, Rossendale

Professor Collette Clifford, Professor of Nursing & Head of Research, School of Health Sciences, University of Birmingham, Edgbaston, Birmingham

Professor Barry Cookson, Director, Laboratory of Healthcare Associated Infection, Health Protection Agency, London

Professor Howard Cuckle, Professor of Reproductive Epidemiology, Department of Paediatrics, Obstetrics & Gynaecology, University of Leeds

Dr Katherine Darton, Information Unit, MIND – The Mental Health Charity, London

Professor Carol Dezateux, Professor of Paediatric Epidemiology, London

Dr Keith Dodd, Consultant Paediatrician, Derby

Mr John Dunning, Consultant Cardiothoracic Surgeon, Cardiothoracic Surgical Unit, Papworth Hospital NHS Trust, Cambridge

Mr Jonothan Earnshaw, Consultant Vascular Surgeon, Gloucestershire Royal Hospital, Gloucester

Professor Martin Eccles, Professor of Clinical Effectiveness, Centre for Health Services Research, University of Newcastle upon Tyne

Professor Pam Enderby, Professor of Community Rehabilitation, Institute of General Practice and Primary Care, University of Sheffield

Mr Leonard R Fenwick, Chief Executive, Newcastle upon Tyne Hospitals NHS Trust

Mrs Gillian Fletcher, Antenatal Teacher & Tutor and President, National Childbirth Trust, Henfield

Professor Jayne Franklyn, Professor of Medicine, Department of Medicine, University of Birmingham, Queen Elizabeth Hospital, Edgbaston, Birmingham

Dr Neville Goodman, Consultant Anaesthetist, Southmead Hospital, Bristol

Professor Robert E Hawkins, CRC Professor and Director of Medical Oncology, Christie CRC Research Centre, Christie Hospital NHS Trust, Manchester

Professor Allen Hutchinson, Director of Public Health & Deputy Dean of ScHARR, Department of Public Health, University of Sheffield

Professor Peter Jones, Professor of Psychiatry, University of Cambridge, Cambridge

Professor Stan Kaye, Cancer Research UK Professor of Medical Oncology, Section of Medicine, Royal Marsden Hospital & Institute of Cancer Research, Surrey

Dr Duncan Keeley, General Practitioner (Dr Burch & Ptnrs), The Health Centre, Thame

Dr Donna Lamping, Research Degrees Programme Director & Reader in Psychology, Health Services Research Unit, London School of Hygiene and Tropical Medicine, London

Mr George Levvy, Chief Executive, Motor Neurone Disease Association, Northampton

Professor James Lindesay, Professor of Psychiatry for the Elderly, University of Leicester, Leicester General Hospital

Professor Julian Little, Professor of Human Genome Epidemiology, Department of Epidemiology & Community Medicine, University of Ottawa

Professor Rajan Madhok, Medical Director & Director of Public Health, Directorate of Clinical Strategy & Public Health, North & East Yorkshire & Northern Lincolnshire Health Authority, York

Professor Alexander Markham, Director, Molecular Medicine Unit, St James's University Hospital, Leeds

Professor Alistair McGuire, Professor of Health Economics, London School of Economics

Dr Peter Moore, Freelance Science Writer, Ashtead

Dr Andrew Mortimore, Consultant in Public Health Medicine, Southampton City Primary Care Trust, Southampton

Dr Sue Moss, Associate Director, Cancer Screening Evaluation Unit, Institute of Cancer Research, Sutton

Mrs Julietta Patnick, Director, NHS Cancer Screening Programmes, Sheffield

Professor Robert Peveler, Professor of Liaison Psychiatry, Royal South Hants Hospital, Southampton

Professor Chris Price, Visiting Chair – Oxford, Clinical Research, Bayer Diagnostics Europe, Cirencester

Professor William Rosenberg, Professor of Hepatology and Consultant Physician, University of Southampton, Southampton

Professor Peter Sandercock, Professor of Medical Neurology, Department of Clinical Neurosciences, University of Edinburgh

Dr Susan Schonfield, Consultant in Public Health, Hillingdon PCT, Middlesex

Dr Eamonn Sheridan, Consultant in Clinical Genetics, Genetics Department, St James's University Hospital, Leeds

Professor Sarah Stewart-Brown, Professor of Public Health, University of Warwick, Division of Health in the Community Warwick Medical School, LWMS, Coventry

Professor Ala Szczepura, Professor of Health Service Research, Centre for Health Services Studies, University of Warwick

Dr Ross Taylor, Senior Lecturer, Department of General Practice and Primary Care, University of Aberdeen

Mrs Joan Webster, Consumer member, HTA – Expert Advisory Network

Feedback

The HTA Programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (http://www.hta.ac.uk) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 23 8059 5639 Email: hta@hta.ac.uk
http://www.hta.ac.uk

ISSN 1366-5278