A systematic review of commercial serological antibody detection tests for the diagnosis of extrapulmonary tuberculosis

Karen R Steingart, Megan Henry, Suman Laal, Philip C Hopewell, Andrew Ramsay, Dick Menzies, Jane Cunningham, Karin Weldingh, Madhukar Pai

Conventional diagnostic tests for tuberculosis have several limitations and are often unhelpful in establishing the diagnosis of extrapulmonary tuberculosis. Although commercial serological antibody based tests are available, their usefulness in the diagnosis of extrapulmonary tuberculosis is unknown. A systematic review was conducted to assess the accuracy of commercial serological antibody detection tests for the diagnosis of extrapulmonary tuberculosis. In a comprehensive search, 21 studies that reported data on sensitivity and specificity for extrapulmonary tuberculosis were identified. These studies evaluated seven different commercial tests, with Anda-TB IgG accounting for 48% of the studies. The results showed that (1) all commercial tests provided highly variable estimates of sensitivity (range 0.00–1.00) and specificity (range 0.59–1.00) for all extrapulmonary sites combined; (2) the Anda-TB IgG kit showed highly variable sensitivity (range 0.26–1.00) and specificity (range 0.59–1.00) for all extrapulmonary sites combined; (3) for all tests combined, sensitivity estimates for both lymph node tuberculosis (range 0.23–1.00) and pleural tuberculosis (range 0.26–0.59) were poor and inconsistent; and (4) there were no data to determine the accuracy of the tests in children or in patients with HIV infection, the two groups for which the test would be most useful. At present, commercial antibody detection tests for extrapulmonary tuberculosis have no role in clinical care or case detection.

Although tuberculosis most commonly affects the lungs, any organ or tissue may be involved. In the USA about 20% of incident cases in 2005 had only extrapulmonary sites of disease and an additional 9% had both pulmonary and extrapulmonary involvement.1 Globally, the proportion of extrapulmonary cases reported by countries ranges from 15% to 25% with greater proportions occurring in countries with a high prevalence of HIV infection.2 In addition to being proportionately greater in persons with HIV infection,3 extrapulmonary involvement occurs with greater relative frequency in children than in adults.4 In children and in persons with HIV infection, extrapulmonary tuberculosis compounds the diagnostic difficulty imposed by their having a lower frequency of sputum smear positivity, even when the lungs are involved.5 6

The diagnosis of extrapulmonary tuberculosis is often difficult to establish, especially for patients in resource limited areas. Signs and symptoms are non-specific and microscopic examination for acid-fast bacilli, the cornerstone of diagnosis for pulmonary tuberculosis in most parts of the world, lacks sensitivity for extrapulmonary disease.10 11 Mycobacterial culture and histological examination for caseating granulomas are more sensitive but not commonly available. Invasive procedures that are complex and costly may be required to obtain the necessary diagnostic specimens.11 12 In a retrospective study of patients in Tanzania with extrapulmonary tuberculosis, bacteriological or histological confirmation of diagnosis was found in only 18%.13 Because of these difficulties, misdiagnosis of extrapulmonary tuberculosis is common in all countries and may result in unnecessary treatment if falsely diagnosed, or greater morbidity and mortality if the diagnosis is missed, especially in persons with HIV infection.14–16

Immune based tests would seem to offer the potential to improve the diagnosis of extrapulmonary tuberculosis as some of the test formats (eg, immunochromatographic test) are practical for resource limited areas. Blood or urine based assays avoid the problems of obtaining a specimen of the affected organ for microbiological or histological assay, are simpler to perform than smear microscopy and the results can be available within hours.4 17 Efforts to develop immune based tests for the detection of antibodies, antigens and immune complexes have been underway for decades and their performance described in several reviews and textbook chapters.16–27 The most common of these tests concentrate on the detection of the humoral (serological) antibody immune response to Mycobacterium tuberculosis (the subject of this review), as opposed to the T cell based cellular immune response (eg, interferon-gamma release assays) or direct detection of antigens in specimens other than serum (eg, lipoarabinomannan detection in urine28 29). It is tempting to speculate that a combination of both humoral and T cell based diagnostic tests could provide the

Abbreviations: AUC, area under the curve; FPR, false positive rate; TPR, true positive rate; ROC, receiver operating characteristic; SROC, summary receiver operating characteristic
highest diagnostic efficacy, although this has not been evaluated to date.

A number of in-house antibody detection tests have been developed but are not marketed. These tests use different antigens and distinct protocols and techniques.

Currently, dozens of commercial serological antibody detection tests (hereafter referred to as commercial tests) are marketed in low income countries where diagnostic tests are rarely subjected to regulatory review or approval.10,11 The extent of their use is unknown; however, companies report sales volumes between 3000 and 30,000 tests per year.12 These tests differ in several respects, including antigen composition and source (eg, native or recombinant), chemical composition (eg, protein, carbohydrate or lipid), extent and manner of antigen purification, class of immunoglobulin (eg, IgG, IgA or IgM), and test format (eg, enzyme-linked immunosorbent assay (ELISA) and immunochromatographic test). Most of the studies investigating the use of antibody detection tests have focused on pulmonary tuberculosis; only a subset has also included patients with extrapulmonary tuberculosis.

To our knowledge, the body of literature evaluating commercial tests for the diagnosis of extrapulmonary tuberculosis has not been synthesised. We therefore conducted a systematic review to summarise the evidence on accuracy (sensitivity and specificity) of commercial tests according to the guidelines and methods proposed for diagnostic systematic reviews and meta-analyses.13 We specifically addressed the following questions:

- Overall, how accurate are commercial tests for the diagnosis of extrapulmonary tuberculosis?
- How accurate are commercial tests for the diagnosis of a specific form of extrapulmonary tuberculosis?

METHODS

Search strategy

The following electronic databases were searched for primary studies in the English language: PubMed (1990 to December 2006), BIOSIS (1990 to December 2005), Embase (1990 to December 2005) and Web of Science (1990 to December 2005). The search terms used included "tuberculosis", "Mycobacterium tuberculosis", "immunological tests", "serological tests", "antibody detection", "antigen detection", "ELISA", "western blot" and "sensitivity and specificity". Additional studies were identified by contacting experts in the field and by searching reference lists from primary studies, review articles and textbook chapters.

Study selection

Extrapulmonary tuberculosis was defined as tuberculosis in which the major site of involvement was outside the lungs. Thereafter, nine types of extrapulmonary disease were classified: (1) eye, (2) pleural, (3) meningeal and/or central nervous system (CNS), (4) bone and/or joint, (5) disseminated/miliary, (6) genitourinary, (7) abdominal, (8) skin and other sites. We included only those studies that based the diagnosis of extrapulmonary tuberculosis on (1) isolation of M tuberculosis on culture or, for studies conducted without culture in tuberculosis endemic countries, the presence of acid-fast bacilli detected by smear microscopy; or (2) the presence of caseating granulomas in histopathological specimens.

Studies that relied solely on clinical and/or radiological features or improvement while on antituberculosis treatment as evidence of tuberculosis were excluded. We further excluded: (1) studies in which the diagnostic criteria were excluded. We further excluded: (1) studies in which the diagnostic criteria were excluded; (2) studies in which (2) the presence of caseating granulomas in histopathological specimens.

Data extraction

One reviewer extracted data on the following qualities: study design, methodological quality, study population, reference standard, site of involvement, antigen and antibody characteristics, laboratory technique, and sensitivity and specificity. To verify the reproducibility of data extraction, a second reviewer independently extracted data from 24% of the included studies. The inter-rater agreement for sensitivity and specificity estimates was 100%. Data not clearly reported were coded as "not reported". When necessary, we attempted to contact authors for additional information.

Although some authors evaluated test performance using several different types of comparison groups, we preferentially selected only one type of comparison group for each study in the following order: (1) patients in whom extrapulmonary tuberculosis was initially suspected but who were later found to have a disease other than tuberculosis; (2) patients in whom pulmonary tuberculosis was initially suspected but who were later found to have non-tuberculous respiratory disease; (3) patients with a variety of diseases other than tuberculosis (mixed disease); (4) healthy persons from tuberculosis endemic countries; (5) contacts of patients with tuberculosis; (6) healthy persons from tuberculosis endemic countries; (5) contacts of patients with tuberculosis; (6) participants from categories 1–5 combined; and (7) healthy persons from non-endemic countries. We felt this hierarchy prioritised the groups in which the test would be used and provide more clinically relevant results.

Assessment of study quality

The quality of studies was assessed using the following criteria, suggested as important for diagnostic studies:19,20

- Was the commercial test result performed and recorded by technicians who were unaware (blinded) of the results of the reference standard?
- Did the whole sample or a randomly selected subset of the sample receive verification using the reference standard?
- Did the study prospectively recruit consecutive patients suspected of having extrapulmonary tuberculosis?

Data collation and meta-analysis

We used standard methods recommended for meta-analyses of diagnostic test evaluations.21,22 As studies were heterogeneous, particularly with respect to the site of involvement, antigen composition of the tests, antibody class (IgG, IgM, or IgA) and control groups, we first grouped studies by type of commercial test and then further stratified by immunoglobulin class and location of disease. If insufficient data (ie, <25 cases) were provided for a specific disease site, we combined data from several sites into a "multiple site" category with at least 25 cases. To calculate sensitivity and specificity of the commercial
tests, we cross-tabulated each result against the reference standard. Sensitivity refers to the proportion of extrapulmonary tuberculosis cases with a positive result on a specific commercial test; specificity refers to the proportion of tuberculosis negative participants that had negative results on a specific commercial test. In calculations of sensitivity we included studies that used smear positivity or histological characteristics as the reference standard together with studies that used culture.

Data were analysed using SPSS (Version 14.0.1.366)35 and Meta-DiSc software (Version 1.4).37 Sensitivity and specificity estimates were calculated for the commercial tests along with their 95% confidence intervals. In addition, true positive rates (TPR = sensitivity) and false positive rates (FPR = 1 – specificity) were summarised using a summary receiver operating characteristic (SROC) curve. Each data point in the SROC space represents an individual study. The SROC curve is obtained by fitting a regression curve to pairs of TPR and FPR.35

### Table 1: Commercial serological antibody detection tests for the diagnosis of extrapulmonary tuberculosis

<table>
<thead>
<tr>
<th>Name of test (number of studies in review)</th>
<th>Antigen(s)</th>
<th>Source of antigen</th>
<th>Ig class</th>
<th>Laboratory technique</th>
<th>Manufacturer</th>
<th>Address/URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anda-TB* (17)</td>
<td>Antigen 60</td>
<td>Native</td>
<td>IgG, IgA, IgM</td>
<td>ELISA</td>
<td>Anda Biologicals S.A.</td>
<td>Strasbourg, France <a href="http://www.andabiologicals.com">www.andabiologicals.com</a></td>
</tr>
<tr>
<td>AMRAD ICT TB (1)</td>
<td>38 kDa and four proprietary antigens</td>
<td>Recombinant</td>
<td>IgG</td>
<td>Immunochromatographic test card</td>
<td>ICT Diagnostics</td>
<td>Balgowlah, NSW, Australia</td>
</tr>
<tr>
<td>Pathozyme Myco [1]</td>
<td>Lipooarabinomannan (LAM) and 38 kDa</td>
<td>Native LAM and recombinant 38 kDa</td>
<td>IgG</td>
<td>ELISA</td>
<td>Omega Diagnostics</td>
<td>Alloa, UK <a href="http://omegadiagnostics.com">http://omegadiagnostics.com</a></td>
</tr>
<tr>
<td>Pathozyme TB Complex Plus (1)</td>
<td>38 kDa and 16 kDa</td>
<td>Recombinant</td>
<td>IgG</td>
<td>ELISA</td>
<td>Omega Diagnostics</td>
<td>Alloa, UK <a href="http://omegadiagnostics.com">http://omegadiagnostics.com</a></td>
</tr>
<tr>
<td>SEVA TB (1)</td>
<td>31 kDa</td>
<td>Native glycoprotein antigen from culture filtrate of M tuberculosis H37Rv</td>
<td>IgG</td>
<td>ELISA</td>
<td>Jammalal Bajaj Tropical Disease Research Centre, Mahatma Gandhi Institute of Medical Sciences</td>
<td>Sevagram-442 102 (Wardha) M S (India)</td>
</tr>
</tbody>
</table>

Ig, immunoglobulin; ELISA, enzyme-linked immunosorbent assay; kDa, kilodalton.

*Anda-TB tests include: IgG (n=10), IgM (n=5), IgA (n=1), IgG and IgM (n=1).
The SROC curve and the area under the curve (AUC) present an overall summary of test performance and display the trade-off between sensitivity and specificity. An AUC of 1.0 indicates perfect discriminatory ability of the diagnostic test. In addition, the Q* index is another useful global summary of the SROC curve and test performance. The Q* index, defined by the point where sensitivity equals specificity on the SROC curve, is the perfect discriminatory ability of the diagnostic test. In addition, heterogeneity could be a result of variability in thresholds, laboratory technique, disease spectrum, study design and/or quality between studies. In the presence of significant heterogeneity, pooled or summary estimates from meta-analyses are difficult to interpret. Given the anticipated variability in accuracy, we decided a priori to avoid the pooling of sensitivity and specificity. Also, as described previously, we addressed heterogeneity by using subgroup (stratified) analyses.

RESULTS
Description of included studies
Of the 3720 citations identified in the literature search, 9 publications describing the results of 21 independent studies...
met our eligibility criteria (fig 1).41–49 None of the studies reported the method (eg, consecutive or random) of subject selection. Only one study48 reported blinded interpretation. No studies involved children younger than 15 years or patients with documented HIV infection. Six studies (29%)41 46 48 were performed in HIV negative patients and 15 (71%) in patients in whom HIV status was unknown or not reported.42–45 47 49 The median sample size was 35 cases (interquartile range (IQR) 30–56) and 48 participants without tuberculosis (IQR 37–194).

ELISA was used in 20 studies and immunochromatography in one study.44 All investigators adhered to standard laboratory methods (eg, mean ± 2SD measured in a healthy population and receiver operating characteristic (ROC) curves) for determining the cut off values, as recommended by the manufacturers. Table 1 shows the commercial tests and their respective antigens. Anda-TB IgG was the most frequently studied test (n = 10 studies (48%)). Tables 2 and 3 show design, quality, performance characteristics and disease site.
In this prospective blinded study (56 as seen in online the 48 Fifty-six studies of Steingart, Henry, Laal, et al [69x52]www.thoraxjnl.com the diagnosis of lymph node tuberculosis. Four studies determined the accuracy of commercial tests for a specific form of extrapulmonary tuberculosis

How accurate are commercial tests for the diagnosis of extrapulmonary tuberculosis?

For all 21 studies the sensitivity estimates ranged from 0.00 to 1.00 and specificity estimates from 0.59 to 1.00 (fig 2). Both sensitivity and specificity varied widely among studies using the same commercial test and among studies using different commercial tests. Confidence intervals for the sensitivity and specificity values of individual studies, depicted graphically by horizontal lines in the forest plots, show poor overlap, suggesting the presence of significant heterogeneity. As shown in fig 3, the accuracy of commercial tests was modest, the symmetrical SROC curve showing a trade off between sensitivity and specificity, with much greater variability in sensitivity estimates.

We identified 10 studies that assessed the accuracy of Anda-TB IgG. As seen in table 2 (and in online supplementary fig S1A available at http://thorax.bmj.com/supplemental), sensitivity estimates ranged from 0.26 to 1.00 and specificity estimates from 0.59 to 1.00. The specificity forest plot (online supplementary fig S1B available at http://thorax.bmj.com/supplemental) includes 7 unique studies as 4 of the total 10 studies were conducted with the same comparison population. The results for both sensitivity and specificity among studies by different investigators were highly variable. Individual study results for the other commercial tests in the review are shown in table 3.

How accurate are commercial tests for the diagnosis of a specific form of extrapulmonary tuberculosis?

Four studies determined the accuracy of commercial tests for the diagnosis of lymph node tuberculosis. As seen in online supplementary figs S2A and B (available at http://thorax.bmj.com/supplemental), sensitivity estimates ranged from 0.23 to 1.00 and specificity estimates from 0.59 to 0.95. Four studies determined the accuracy of commercial tests for the diagnosis of pleural tuberculosis. Sensitivity estimates ranged from 0.26 to 0.59 and specificity estimates, from 0.81 to 1.00 (see online supplementary figs S3A and B available at http://thorax.bmj.com/supplemental). For both lymph node and pleural tuberculosis, commercial tests showed inconsistent results.

DISCUSSION

Principal findings

Our systematic review of 21 studies examining the performance of commercial tests for the diagnosis of extrapulmonary tuberculosis showed that (1) all commercial tests provided highly variable estimates of sensitivity (range 0.00–1.00) and specificity (range 0.59–1.00) for all extrapulmonary sites combined; (2) Anda-TB IgG showed highly variable sensitivity (range 0.26–1.00) and specificity (range 0.59–1.00) for all extrapulmonary sites combined; (3) for all commercial tests combined, sensitivity estimates for both lymph node tuberculosis (range 0.23–1.00) and pleural tuberculosis (range 0.26–0.59) were poor and inconsistent; and (4) there were no data to determine the accuracy of commercial tests for the diagnosis of extrapulmonary tuberculosis in children or patients with HIV infection. Although commercial serological tests—by virtue of being rapid, simple to use and non-invasive—are appealing, this review did not find sufficient evidence to justify their use for the diagnosis of extrapulmonary tuberculosis.

Our systematic review had several strengths. First, the comprehensive search strategy with various overlapping approaches enabled us to retrieve relevant studies published since 1990. Moreover, two reviewers independently completed screening and study selection. To verify reproducibility of data extraction, a second reviewer independently extracted data on five (24%) of the included studies. Whenever possible we selected a control population with disease, in lieu of healthy subjects, to evaluate how well commercial tests performed in patients suspected of having tuberculosis. Authors were contacted for missing data. Finally, we analysed data within specific subgroups to lessen the effect of heterogeneity.

This review also had limitations. There were an insufficient number of studies for most of the specific commercial tests or the specific disease sites to provide meaningful summary measures of performance. Our use of stringent criteria for eligibility is probably the main reason that we identified only one study on tuberculous meningitis. Fifty-six studies of tuberculous meningitis identified by our search strategy were considered ineligible because they involved fewer than 25 cases; investigated studies of fluids (eg, cerebrospinal fluid) other than serum; involved antigen detection; or relied on clinical features and/or treatment response for case confirmation. Our choice of a bacteriological and/or histopathological reference standard may have limited the inclusion of studies involving children. Paediatric tuberculosis is difficult to diagnose on a bacteriological basis because of the paucibacillary nature of the disease. In addition, the number of specific antigens included in the commercial tests in this review was limited (table 1) compared with the number of potential antigens for serodiagnosis.

Another set of problems involved shortcomings in study design and quality. No studies reported the method for recruitment of subjects, so it was not possible to ascertain if studies used the sound probabilistic sampling framework found in consecutive or random sampling designs. Only one study reported a blinded interpretation of the results of the commercial test and the reference standard. Lack of blinding may have resulted in an overestimation of the accuracy of the commercial test. Variability in study design and study quality might account for some of the observed heterogeneity evident in the results. Although statistical tests and graphical methods only one study assessed the accuracy of a commercial test in patients with meningitis. In this prospective blinded study (56 culture confirmed patients and 74 controls), the sensitivity of the AMRAD ICT TB test was 0.48 (95% CI 0.35 to 0.62) and the specificity was 0.82 (95% CI 0.72 to 0.90).

www.thoraxjnl.com
Antibody detection tests for extrapulmonary tuberculosis

... available to detect potential publication bias in meta-analyses of randomised control trials, such techniques have not been adequately evaluated for diagnostic data. It is therefore difficult to rule out publication bias in our review. In addition, our search strategy may have missed some relevant studies by excluding non-English language publications.

Developing an immunological diagnostic test for tuberculosis presents a formidable challenge in part because both the stage of tuberculosis infection and the tissues involved may alter the profile of genes expressed by the organism and, thus, the antibody responses to these gene products may differ. Studies during the last decade have provided ample evidence that M tuberculosis adapts to its environment by altering the profile of genes that it expresses, that these profiles are modulated as infection progresses and the in vivo environment changes, and that some genes of M tuberculosis are differentially expressed in different host tissues. Consequently, antibodies developed in response to pulmonary tuberculosis may not be the appropriate targets for diagnosing extrapulmonary involvement. The choice of reagents for all current assays was most likely determined in patients with pulmonary tuberculosis. Although a systematic investigation of the humoral immune responses of pulmonary tuberculosis has been performed and several antigenic proteins that are recognised by antibodies at different stages of pulmonary disease have been identified, no similar analysis of antigens expressed during extrapulmonary replication of M tuberculosis has been attempted so far. Thus, identification and study of M tuberculosis genes expressed in the different environments that characterise different sites of involvement may be able to provide the optimal reagents for devising a diagnostic test for extrapulmonary forms of the disease.

It will also be important to examine proteins expressed by M tuberculosis in HIV infected patients as smear negative pulmonary and extrapulmonary disease are disproportionately higher in HIV positive than in HIV negative individuals. Given that memory B cells are relatively independent of T cell help, antibody detection based diagnostic tests would be significant assets for identification of forms of paucibacillary disease. Indeed, despite the dysfunctional cellular immune responses and the presence of HIV induced hypergammaglobinaemia, the presence of antibodies to M tuberculosis antigens has been reported by several investigators. It is also possible that antibodies to M tuberculosis antigens are elicited in patients who get infected before their immune system is significantly compromised and the CD4 numbers are still high, but we are unaware of studies that have addressed this possibility.

CONCLUSIONS AND POLICY IMPLICATIONS

The evidence presented in this systematic review shows that, at present, commercial tests produce highly variable sensitivity and specificity results and therefore cannot be recommended as a sole test for the diagnosis of extrapulmonary tuberculosis. It is particularly disappointing that there are no studies of commercial tests that are of sufficient quality to enable their evaluation in patients with HIV infection or in children, as it is in these groups that the tests could be most useful. Our findings should be interpreted in the context of the variability in design and the quality of the studies in this review. Recent articles have attested to the mediocre quality of diagnostic studies for tuberculosis. The use of guidelines such as the Standards for Reporting of Diagnostic Accuracy (STARD) and the tool for Quality Assessment of Diagnostic Accuracy Studies (QUADAS) may lead to improvements in the quality of future studies. Guidelines specifically for the evaluation of diagnostic tests for infectious diseases have recently been published.

ACKNOWLEDGEMENTS

The authors thank Nandini Dedukuri, Maya Bhat-Gregerson and Anna Meddaugh for technical assistance, Madelyn Hall for help in acquiring the publications in the review, and Izabela Suder-Dayao and Melissa Anthony of WHO/TDR for administrative assistance.

Further data are given in the online figures available at http://thorax.bmj.com/supplemental.

Authors' affiliations

Kate A Steingart, Philip C Hopewell, Division of Pulmonary and Critical Care Medicine, San Francisco General Hospital, University of California, and Francis J Curry National Tuberculosis Center, San Francisco, California, USA

Megan Henry, County of Sacramento Department of Health and Human Services, Sacramento, California, USA

Suman Laal, Veterans Affairs Medical Center and Departments of Pathology and Microbiology, New York University School of Medicine, New York, USA

Andrew Ramsay, Jane Cunningham, UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR), World Health Organization, Geneva, Switzerland

Dick Menzies, Madhukar Pai, Respiratory Epidemiology and Clinical Research Unit, Montreal Chest Institute and Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada

Karim Weldingh, Statens Serum Institut, Department of Infectious Disease Immunology, Copenhagen S, Denmark

This work was supported by the UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR). AR and JC, both with TDR, contributed to the conception and design of the systematic review, and critical revision and decision to publish the manuscript. AR also participated in data interpretation. Competing interests: KW is co-inventor on a number of patents relating to mycobacterial antigens which may be used for serological assays. All rights have been assigned to Statens Serum Institut.

REFERENCES


