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

Karen R. Steingart\textsuperscript{1,2}, Megan Henry\textsuperscript{3}, Suman Laal\textsuperscript{4,5,6}, Philip C. Hopewell\textsuperscript{1,2}, Andrew Ramsay\textsuperscript{7}, Dick Menzies\textsuperscript{8}, Jane Cunningham\textsuperscript{7}, Karin Weldingh\textsuperscript{9}, Madhukar Pai\textsuperscript{8,10}

1 Division of Pulmonary and Critical Care Medicine, San Francisco General Hospital, University of California, San Francisco, California, United States of America
2 Francis J. Curry National Tuberculosis Center, San Francisco, California, United States of America
3 San Joaquin County Public Health Services, Stockton, California, United States of America
4 Department of Pathology, New York University School of Medicine, New York, New York, United States of America
5 Department of Microbiology, New York University School of Medicine, New York, New York, United States of America
6 Department of Veterans Affairs, New York Harbor Healthcare System, New York, New York, United States of America
7 UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR), World Health Organization, Geneva, Switzerland
8 Respiratory Epidemiology & Clinical Research Unit, Montréal Chest Institute, Montréal, Québec, Canada
9 Statens Serum Institut
Department of Infectious Disease Immunology, Copenhagen S, Denmark
10 Department of Epidemiology, Biostatistics & Occupational Health, McGill University, Montréal, Québec, Canada

Keywords: Tuberculosis, Extrapulmonary/*diagnosis; Immunoassay; Serologic Tests/methods; Sensitivity and Specificity

*Correspondence to:
Dr Madhukar Pai
Department of Epidemiology, Biostatistics & Occupational Health
McGill University, 1020 Pine Avenue West, Montréal, Québec, H3A 1A2, Canada
madhukar.pai@mcgill.ca
Abstract
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 utility in the diagnosis of extrapulmonary tuberculosis is unknown. We conducted a systematic review to assess the accuracy of commercial serological antibody detection tests for the diagnosis of extrapulmonary tuberculosis. In a comprehensive search, we identified 21 studies that reported data on sensitivity and specificity for extrapulmonary tuberculosis. These studies evaluated five different commercial tests, with Anda-TB IgG accounting for 48% of the studies. Results of the review demonstrated that (1) all commercial tests provided highly variable estimates of sensitivity (range 0% to 100%) and specificity (range 59% to 100%) for all extrapulmonary sites combined; (2) the Anda-TB IgG kit showed highly variable sensitivity (range 26% to 100%) and specificity (range 59% to 100%) for all extrapulmonary sites combined; (3) for all tests combined, sensitivity estimates for both lymph node tuberculosis (range 23% to 100%) and pleural tuberculosis (range 26% to 59%) were poor and inconsistent; and (4) there were no data to determine accuracy of the tests in children or in patients with HIV infection, the two groups for which the test would be most useful. At the current time, commercial antibody detection tests for extrapulmonary tuberculosis have no role in clinical care or case detection.
Introduction

Although tuberculosis most commonly affects the lungs, any organ or tissue may be involved. In the United States about 20% of incident cases in 2005 had only extrapulmonary sites of disease and an additional 9% had both pulmonary and extrapulmonary involvement. Globally, the proportion of extrapulmonary tuberculosis cases is approximately 15% to 25% of tuberculosis cases with greater proportions occurring in countries with a high prevalence of HIV infection. In addition to being proportionately greater in persons with HIV infection, extrapulmonary involvement occurs with greater relative frequency in children compared with adults. 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.

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. 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. In a retrospective study of patients in Tanzania with extrapulmonary tuberculosis, bacteriological or histological confirmation of diagnosis was found in only 18%. 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.

Immune-based tests would seem to offer the potential to improve the diagnosis of extrapulmonary tuberculosis, as some of the test formats (e.g., 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 results can be available within hours. 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. The most common of these tests concentrate on the detection of the humoral (serological) antibody immune response to \textit{Mycobacterium tuberculosis} (\textit{M. tuberculosis}) (the subject of this review), as opposed to the T-cell based cellular immune response (e.g., interferon-gamma release assays), or direct detection of antigens in specimens other than serum (e.g., lipoarabinomannan [LAM] detection in urine). It is tempting to speculate that a combination of both humoral and T-cell based diagnostic tests could provide the highest diagnostic efficacy, although this has not been evaluated so far.

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. The extent of their use is unknown; however,
companies report sales volumes between 3,000 and 300,000 tests per year. These tests differ in several respects, including antigen composition and source (e.g., native or recombinant), chemical composition (e.g., protein, carbohydrate, or lipid), extent and manner of antigen purification, class of immunoglobulin (e.g., IgG, IgA, or IgM), and test format (e.g., enzyme-linked immunosorbent assay [ELISA] and immunochromatographic test). A majority of 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 synthesized. We, therefore, conducted a systematic review to summarize the evidence on accuracy (sensitivity and specificity) of commercial tests according to the guidelines and methods proposed for diagnostic systematic reviews and meta-analyses. We specifically addressed the questions: (1) Overall, how accurate are commercial tests for the diagnosis of extrapulmonary tuberculosis? and (2) How accurate are commercial tests for the diagnosis of a specific form of extrapulmonary tuberculosis?

Methods

Search strategy


Study selection

We defined extrapulmonary tuberculosis as tuberculosis in which the major site of involvement was outside the lungs. Thereafter, nine types of extrapulmonary disease were classified: lymph node, pleural, meninges and/or central nervous system (CNS), bone and/or joint, disseminated/miliary, genitourinary, abdominal, 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.

We excluded studies that relied solely on clinical and/or radiologic features or improvement while on antituberculosis treatment as the diagnostic criteria. We further excluded: (1) studies published before 1990 for the reason that many studies used crude extracts or obsolete immunologic methods; (2) studies with fewer than 50 participants (at least 25 tuberculosis patients and 25 participants without tuberculosis were required for inclusion); (3) studies in which data were only provided for pulmonary and extrapulmonary cases combined; (4) studies of fluids other than serum (e.g., cerebrospinal fluid); (5) studies of latent tuberculosis infection; (6) studies focused on nontuberculous mycobacteria; (7) studies of antibody responses during
or after tuberculosis treatment; (8) investigations using non-immunologic methods for detection of antibodies; (9) basic science literature that focused on cloning of new antigens or their immunologic properties (e.g., epitope mapping) or other new methods of antibody detection; and (10) case reports and reviews.

Initially, two reviewers (KRS and MH) screened citations retrieved from all sources. To identify eligible studies, a second screening was done (KRS and MH) of full texts from citations found relevant in the first screen. A list of excluded studies and reasons for their exclusion is available from the authors.

Data extraction
One reviewer (KRS) 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 reproducibility of data extraction, a second reviewer (MH) 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 compared performance of commercial tests in several different groups without TB, we preferentially selected only one 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 nontuberculous 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) participants from categories (1) through (5) combined; and (7) healthy persons from non-endemic countries. We felt this hierarchy prioritized the populations in which the test would be used and provided more clinically relevant results.

Assessment of study quality
We assessed the quality of studies using the following criteria, suggested as important for diagnostic studies:33 34 (1) Was the commercial test result performed and recorded by technicians who were unaware (blinded) of the results of the reference standard? (2) Did the whole sample or a randomly selected subset of the sample receive verification using the reference standard? (3) 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.33 35 As studies were heterogeneous, particularly with respect to the site of involvement, antigen composition of the tests, antibody class (IgG, IgM, or IgA), and comparison groups, we first grouped studies by type of commercial test and then further stratified by immunoglobulin class and location of disease. If insufficient data (i.e., fewer than 25 tuberculosis patients) were provided for a specific disease site, we combined data from several sites into a “multiple site” category with at least 25 participants. 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 patients with a positive result on a specific commercial test; specificity refers to the proportion of participants without tuberculosis 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 along with studies that used culture.

Data were analyzed using SPSS (version 14.0.1.366) and Meta-DiSc software (version 1.4). 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 summarized 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.

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 1.00 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 point that is intersected by the anti-diagonal, the top-left corner of the SROC region. A Q* value of 1.00 indicates 100% accuracy.

In meta-analyses of studies of diagnostic tests, heterogeneity refers to a high level of variability in study results. Such 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, nine publications describing the results of 21 independent studies met our eligibility criteria (fig 1). None of the studies reported the method (e.g., consecutive or random) of subject selection. Only one (5%) study reported blinded interpretation. No studies involved children younger than 15 years of age or patients with documented HIV infection. Six (29%) studies were performed in HIV-negative patients and 15 (71%) studies in patients in whom HIV status was unknown or not reported. The median number of tuberculosis patients was 35 (interquartile range 30 to 56); the median number of participants without tuberculosis was 48 (interquartile range 37 to 194).

ELISA was used in 20 studies and immunochromatographic test in one study. All investigators adhered to standard laboratory methods (e.g., mean +/- 2 SD measured in a healthy population and receiver operating characteristic (ROC) curves) for determining the cutoff values, as recommended by the manufacturers. Five different commercial tests (one test
used distinct assays for the detection of IgG, IgM, and IgA antibodies) were identified with Anda-TB IgG being the most frequently studied test (10 [48%] studies) (table 1).
**Table 1** Commercial serological antibody detection tests for the diagnosis of extrapulmonary tuberculosis

<table>
<thead>
<tr>
<th>Name of Test (Number of Studies in the Review)</th>
<th>Antigen(s)</th>
<th>Antigen Source</th>
<th>Immunoglobulin Class</th>
<th>Laboratory Technique</th>
<th>Name of 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>ICT TB (1)</td>
<td>38kDa and four proprietary antigens</td>
<td>Recombinant</td>
<td>IgG</td>
<td>Immunochromatographic test card ELISA</td>
<td>ICT Diagnostics</td>
<td>Balgowlah, New South Wales, Australia (<a href="http://www.ictdiagnostics.com/">www.ictdiagnostics.com/</a>)</td>
</tr>
<tr>
<td>Pathozyme Myco (1)</td>
<td>LAM and 38kDa</td>
<td>Native LAM and recombinant 38 kDa</td>
<td>IgG</td>
<td>ELISA</td>
<td>Omega Diagnostics</td>
<td>Alloa, Scotland (<a href="http://www.omegadiagnostics.com/">http://www.omegadiagnostics.com/</a>)</td>
</tr>
<tr>
<td>Pathozyme TB Complex Plus (1)</td>
<td>38kDa and 16kDa</td>
<td>Recombinant</td>
<td>IgG</td>
<td>ELISA</td>
<td>Omega Diagnostics</td>
<td>Alloa, Scotland (<a href="http://www.omegadiagnostics.com/">http://www.omegadiagnostics.com/</a>)</td>
</tr>
<tr>
<td>SEVA TB (1)</td>
<td>31 kDa</td>
<td>Native glycoprotein antigen from culture filtrate of <em>M. tuberculosis</em> H37Rv</td>
<td>IgG</td>
<td>ELISA</td>
<td>Jamnalal Bajaj Tropical Disease Research Centre Mahatma Gandhi Institute of Medical Sciences</td>
<td>Sevagram-442 102 (Wardha) M. S. India</td>
</tr>
</tbody>
</table>

* Anda-TB tests include the following: IgG (10); IgM (5); IgA (1); IgG and IgM (1) ELISA, enzyme-linked immunosorbent assay; kDa, kilodalton; LAM, lipoarabinomannan
Overall, how accurate are commercial tests for the diagnosis of extrapulmonary tuberculosis?

Tables 2 and 3 show performance and other selected characteristics for the commercial tests in the review. When all 21 studies were considered together, the sensitivity estimates ranged from 0% to 100% and specificity estimates, from 59% to 100% (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 seen in figure 3, the accuracy of commercial tests was modest, the symmetric SROC curve showing a trade off between sensitivity and specificity, with much greater variability in sensitivity estimates.
Table 2 Selected characteristics of studies investigating Anda-TB (Anda Biologicals, Strasbourg, France) for the diagnosis of extrapulmonary tuberculosis

<table>
<thead>
<tr>
<th>Author (Year, Study)</th>
<th>Data Collection</th>
<th>Verification Standard</th>
<th>Disease Site</th>
<th>Disease Site Location</th>
<th>Country</th>
<th>Type of Comparison Group</th>
<th>Immuno globulin Class</th>
<th>No. Participants</th>
<th>Sensitivity (95% Confidence Interval)</th>
<th>Specificity (95% Confidence Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alifano (1998, a)</td>
<td>Retrospective</td>
<td>Differential</td>
<td>Culture and/or histology</td>
<td>Multiple</td>
<td>Italy</td>
<td>Mixed disease</td>
<td>IgG</td>
<td>42/44</td>
<td>0.74 (0.58-0.86)</td>
<td>0.93 (0.81-0.99)</td>
</tr>
<tr>
<td>Alifano (1998, b)</td>
<td>Retrospective</td>
<td>Differential</td>
<td>Culture and/or histology</td>
<td>Multiple</td>
<td>Italy</td>
<td>Mixed disease</td>
<td>IgA</td>
<td>42/44</td>
<td>0.69 (0.53-0.82)</td>
<td>0.89 (0.75-0.96)</td>
</tr>
<tr>
<td>Banerjee (2003, a)</td>
<td>Retrospective</td>
<td>NR</td>
<td>Lymph node</td>
<td>India</td>
<td>Healthy</td>
<td>IgG</td>
<td>30/32</td>
<td>0.43 (0.25-0.63)</td>
<td>0.59 (0.41-0.76)</td>
<td></td>
</tr>
<tr>
<td>Caminero (1993)</td>
<td>Prospective</td>
<td>Complete</td>
<td>Culture and/or histology</td>
<td>Pleura</td>
<td>Spain</td>
<td>Pleural TB suspects</td>
<td>IgG</td>
<td>30/48</td>
<td>0.53 (0.34-0.72)</td>
<td>1.00 (0.93-1.00)</td>
</tr>
<tr>
<td>Caminero (1994)</td>
<td>Prospective</td>
<td>Differential</td>
<td>Culture and/or histology</td>
<td>Multiple</td>
<td>Canary Islands</td>
<td>Nontuberculous respiratory disease</td>
<td>IgG</td>
<td>56/31</td>
<td>0.32 (0.20-0.46)</td>
<td>0.94 (0.79-0.99)</td>
</tr>
<tr>
<td>Gevaudan (1992, a)</td>
<td>Retrospective</td>
<td>Differential</td>
<td>Culture and/or histology</td>
<td>Lymph node</td>
<td>France</td>
<td>Mixed disease</td>
<td>IgG</td>
<td>26/194</td>
<td>1.00 (0.87-1.00)</td>
<td>0.76 (0.69-0.82)</td>
</tr>
<tr>
<td>Gevaudan (1992, b)</td>
<td>Retrospective</td>
<td>Differential</td>
<td>Culture and/or histology</td>
<td>Lymph node</td>
<td>France</td>
<td>Mixed disease</td>
<td>IgM</td>
<td>26/194</td>
<td>0.23 (0.09-0.44)</td>
<td>0.95 (0.91-0.98)</td>
</tr>
<tr>
<td>Gevaudan (1992, c)</td>
<td>Retrospective</td>
<td>Differential</td>
<td>Culture</td>
<td>Disseminated/ miliary (primary)</td>
<td>France</td>
<td>Mixed disease</td>
<td>IgG</td>
<td>56/194</td>
<td>0.95 (0.85-0.99)</td>
<td>0.76 (0.69-0.82)</td>
</tr>
<tr>
<td>Gevaudan (1992, d)</td>
<td>Retrospective</td>
<td>Differential</td>
<td>Culture</td>
<td>Disseminated/ miliary (primary)</td>
<td>France</td>
<td>Mixed disease</td>
<td>IgM</td>
<td>56/194</td>
<td>0.32 (0.20-0.46)</td>
<td>0.95 (0.91-0.98)</td>
</tr>
<tr>
<td>Gevaudan (1992, e)</td>
<td>Retrospective</td>
<td>Differential</td>
<td>Culture</td>
<td>Disseminated/ miliary (post-primary)</td>
<td>France</td>
<td>Mixed disease</td>
<td>IgG</td>
<td>25/194</td>
<td>1.00 (0.86-1.00)</td>
<td>0.76 (0.69-0.82)</td>
</tr>
<tr>
<td>Gevaudan (1992, f)</td>
<td>Retrospective</td>
<td>Differential</td>
<td>Culture</td>
<td>Disseminated/ miliary</td>
<td>France</td>
<td>Mixed disease</td>
<td>IgM</td>
<td>25/194</td>
<td>0.00 (0.00-0.14)</td>
<td>0.95 (0.91-0.98)</td>
</tr>
</tbody>
</table>
Antibody detection tests for extrapulmonary tuberculosis

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Comparisons</th>
<th>Location</th>
<th>Disease</th>
<th>Marker</th>
<th>Cases</th>
<th>Controls</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gevaudan (1992, g)</td>
<td>Retrospective</td>
<td>Differential Culture (post-primary) Genitourinary</td>
<td>France</td>
<td>Mixed disease</td>
<td>IgG</td>
<td>34/194</td>
<td>0.76 (0.69-0.82)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gevaudan (1992, h)</td>
<td>Retrospective</td>
<td>Differential Culture Genitourinary</td>
<td>France</td>
<td>Mixed disease</td>
<td>IgM</td>
<td>34/194</td>
<td>0.95 (0.91-0.98)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kunter (2003, a)</td>
<td>Prospective</td>
<td>Differential Culture Pleura</td>
<td>Turkey</td>
<td>Nontuberculous respiratory disease</td>
<td>IgG</td>
<td>88/37</td>
<td>0.86 (0.71-0.96)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kunter (2003, b)</td>
<td>Prospective</td>
<td>Differential Culture Pleura</td>
<td>Turkey</td>
<td>Nontuberculous respiratory disease</td>
<td>IgM</td>
<td>88/37</td>
<td>0.92 (0.78-0.98)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kunter (2003, c)</td>
<td>Prospective</td>
<td>Differential Culture Pleura</td>
<td>Turkey</td>
<td>Nontuberculous respiratory disease</td>
<td>IgG and IgM</td>
<td>88/37</td>
<td>0.81 (0.65-0.92)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luh (1996)</td>
<td>Prospective</td>
<td>NR Culture</td>
<td>Multiple Taiwan</td>
<td>Mixed disease</td>
<td>IgG</td>
<td>35/224</td>
<td>0.92 (0.88-0.96)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: The treatment status of cases was not reported for reference 45. For all other studies, patients were either not on treatment or on treatment for less than 14 days at the time serum was obtained.

*Number of participants with tuberculosis/without tuberculosis

IgG, IgM, IgA, immunoglobulin G, M, A, respectively; NR, not reported.
Table 3 Selected characteristics of studies investigating commercial tests for the diagnosis of extrapulmonary tuberculosis

<table>
<thead>
<tr>
<th>Author (Year, Study)</th>
<th>Data Collection</th>
<th>Verification</th>
<th>Reference Standard</th>
<th>Disease Site</th>
<th>Country</th>
<th>Type of Comparison Group</th>
<th>Name of Test&lt;sup&gt;a&lt;/sup&gt;</th>
<th>No. Participants&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Sensitivity (95% Confidence Interval)</th>
<th>Specificity (95% Confidence Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Banerjee (2003, b)&lt;sup&gt;42&lt;/sup&gt;</td>
<td>Retrospective</td>
<td>NR</td>
<td>Histology</td>
<td>Lymph node</td>
<td>India</td>
<td>Healthy</td>
<td>SEVA TB</td>
<td>30/32</td>
<td>0.77 (0.58-0.90)</td>
<td>0.88 (0.71-0.97)</td>
</tr>
<tr>
<td>McConkey (2002)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Prospective</td>
<td>Complete</td>
<td>Culture</td>
<td>Meninges</td>
<td>Egypt</td>
<td>Mixed disease</td>
<td>ICT TB</td>
<td>56/74</td>
<td>0.48 (0.35-0.62)</td>
<td>0.82 (0.72-0.90)</td>
</tr>
<tr>
<td>Nsanze (1997, a)&lt;sup&gt;49&lt;/sup&gt;</td>
<td>Prospective</td>
<td>Differential</td>
<td>Culture and/or smear</td>
<td>NR</td>
<td>The United Arab Emirates</td>
<td>Mixed disease and healthy</td>
<td>Pathozyme TB Complex Plus</td>
<td>35/35</td>
<td>0.51 (0.34-0.69)</td>
<td>1.00 (0.90-1.00)</td>
</tr>
<tr>
<td>Nsanze (1997, b)&lt;sup&gt;49&lt;/sup&gt;</td>
<td>Prospective</td>
<td>Differential</td>
<td>Culture and/or smear</td>
<td>NR</td>
<td>The United Arab Emirates</td>
<td>Mixed disease and healthy</td>
<td>Pathozyme Myco</td>
<td>35/35</td>
<td>0.11 (0.03-0.27)</td>
<td>1.00 (0.90-1.00)</td>
</tr>
</tbody>
</table>

<sup>a</sup>All commercial tests detected immunoglobulin G antibody

<sup>b</sup>Number of participants with tuberculosis/without tuberculosis

<sup>c</sup>Blinded study

NR, not reported
We identified ten studies that assessed the accuracy of Anda-TB IgG.\textsuperscript{41-47} As seen in table 2 (and in online supplementary fig S1A), sensitivity estimates ranged from 26% to 100% and specificity estimates, from 59% to 100%. The specificity forest plot (online supplementary fig S1B) includes seven unique studies, as four\textsuperscript{45} of the total ten studies were conducted with the same comparison group. 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.\textsuperscript{42,45} As seen in online supplementary fig S2A, sensitivity estimates ranged from 23% to 100% and specificity estimates, from 59% to 95% (online supplementary fig S2B). Four studies determined the accuracy of commercial tests for the diagnosis of pleural tuberculosis.\textsuperscript{43,46} Sensitivity estimates ranged from 26% to 59% and specificity estimates, from 81% to 100% (online supplementary figs S3A and B). For both lymph node and pleural tuberculosis, commercial tests showed inconsistent results.

Only one study assessed the accuracy of a commercial test in patients with meningitis.\textsuperscript{48} In this prospective, blinded study (56 culture-confirmed tuberculosis patients and 74 participants without tuberculosis), the sensitivity of ICT TB test was 48% (95% CI 35-62%) and specificity, 82% (95% CI 72-90%).

**Discussion**

Principal findings

Our systematic review of 21 studies examining the performance of commercial tests for the diagnosis of extrapulmonary tuberculosis demonstrated that (1) all commercial tests provided highly variable estimates of sensitivity (range 0% to 100%) and specificity (range 59% to 100%) for all extrapulmonary sites combined; (2) Anda-TB IgG showed highly variable sensitivity (range 26% to 100%) and specificity (range 59% to 100%) for all extrapulmonary sites combined; (3) for all commercial tests combined, sensitivity estimates for both lymph node tuberculosis (range 23% to 100%) and pleural tuberculosis (range 26% to 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 comparison group with disease, in lieu of healthy participants, to evaluate how well commercial tests performed in patients suspected of having tuberculosis. We contacted authors for missing data. Finally, we analyzed data within specific subgroups to lessen the effect of heterogeneity.
This review also had limitations. There were an insufficient number of studies for the majority of either the specific commercial test or the specific disease site to provide meaningful summary measures of performance. Our use of stringent criteria for eligibility is likely 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 tuberculosis patients; investigated studies of fluids (e.g., cerebrospinal) other than serum; involved antigen detection; or relied on clinical features and/or treatment response for case confirmation. Our choice of a bacteriologic and/or histopathological reference standard may have limited the inclusion of studies involving children. Pediatric tuberculosis is difficult to diagnose on a bacteriological basis because of the paucibacillary nature of disease. In addition, the number of specific antigens included in the commercial tests in this review was limited (table 1) compared to 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 participants; therefore, it was not possible to ascertain if studies used the sound probabilistic sampling framework found in consecutive or random sampling designs. Only one (5%) study reported 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 are available to detect potential publication bias in meta-analyses of randomized control trials, such techniques have not been adequately evaluated for diagnostic data. Thus, it is 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 immunologic 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 recognized 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 characterize 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 hypergammaglobulinemia, 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 demonstrates that, at the current time, 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. 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.

**Funding**

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 conception and design of the systematic review, and critical revision of and decision to publish the manuscript. AR also participated in data interpretation.

**Competing interests**

KW is coinventor 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.

**Author contributions**

KRS, SL, PCH, AR, DM, JC, KW, and MP conceived and designed the study. KRS and MH collected and analyzed the data. KRS, MH, SL, AR, DM, KW, and MP interpreted the data. KRS, SL, and MP drafted the manuscript. KRS, SL, PCH, AR, DM, JC, KW, and MP performed critical revision of the manuscript.

**Acknowledgments**

We wish to thank Nandini Dendukuri, Maya Bhat-Gregerson, and Anna Meddaugh for technical assistance and Madelyn Hall for help in acquiring the publications in the review.
would like to thank Izabela Suder-Dayao and Melissa Anthony of WHO/TDR for administrative assistance.

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in Thorax editions and any other BMJPGL products to exploit all subsidiary rights, as set out in our licence (http://thorax.bmjjournals.com/ifora/licence.pdf).
**Figure 1** Flow diagram for study selection

(A)

(B)

**Figure 2** Sensitivity and specificity estimates of commercial tests for the diagnosis of extrapulmonary tuberculosis. (A) Sensitivity; (B) Specificity. The circles and lines represent the point estimates and 95% CIs, respectively. The size of the circle indicates the study size.

**Figure 3** Summary receiver operating characteristic (SROC) curve for commercial tests. Each solid circle represents each study in the meta-analysis. The curve is the regression line that summarizes the overall diagnostic accuracy. AUC=area under the curve; SE (AUC)=standard error of AUC; Q*=an index defined by the point on the SROC curve where the sensitivity and specificity are equal, which is the point closest to the top-left corner of the ROC space; SE (Q*)=standard error of Q* index.

(A)

(B)

**Figure S1** Sensitivity and specificity estimates of Anda-TB IgG for the diagnosis of extrapulmonary tuberculosis. (A) Sensitivity; (B) Specificity. The circles and lines represent the point estimates and 95% CIs, respectively. The size of the circle indicates the study size.
**Figure S2** Sensitivity and specificity estimates of commercial tests for diagnosis of lymph node tuberculosis. (A) Sensitivity; (B) Specificity. The circles and lines represent the point estimates and 95% CIs, respectively. The size of the circle indicates the study size. Banerjee 2003 (a)=Anda-TB IgG; Banerjee 2003 (b)=SEVA TB; Gevaudan 1992 (a)=Anda-TB IgG; Gevaudan 1992 (b)=Anda-TB IgM

**Figure S3** Sensitivity and specificity estimates of commercial tests for the diagnosis of pleural tuberculosis. (A) Sensitivity; (B) Specificity. The circles and lines represent the point estimates and 95% CIs, respectively. The size of the circle indicates the study size. Caminero 1993=Anda-TB IgG; Kunter 2003 (a)=Anda-TB IgG; Kunter 2003 (b)=Anda-TB IgM; Kunter 2003 (c)=Anda-TB IgG and IgM combined.
References


3720 potentially relevant citations identified from electronic databases

Excluded after first screen: 3205
Reasons for exclusion:
Duplicates: 298
Non-English language: 746
On basis of title/abstract: 2161

515 articles selected for full-text review

Excluded: 482
Reasons for exclusion:
In-house antibody test: 137
Insufficient no. participants: 70
Not serum antibody detection: 61
Tuberculosis form unspecified: 28
Sputum status not specified: 50
Insufficient case confirmation: 44
Relevance to topic: 34
Abstract/letter/case report: 17
Review/editorial: 16
No test accuracy/methodology: 14
Duplicate publication: 5
Patients on treatment: 4
Animal study: 1
Non-English language: 1

33 articles met eligibility criteria for all forms of tuberculosis

Excluded: Pulmonary tuberculosis: 24

9 articles (21 studies) in systematic review of commercial antibody detection tests for extrapulmonary tuberculosis
Fig S2B

<table>
<thead>
<tr>
<th>Study</th>
<th>Specificity</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Banerjee 2003 (a)</td>
<td>0.59</td>
<td>(0.41 - 0.7)</td>
</tr>
<tr>
<td>Banerjee 2003 (b)</td>
<td>0.88</td>
<td>(0.71 - 0.9)</td>
</tr>
<tr>
<td>Gevaudan 1992 (a)</td>
<td>0.76</td>
<td>(0.69 - 0.8)</td>
</tr>
<tr>
<td>Gevaudan 1992 (b)</td>
<td>0.95</td>
<td>(0.91 - 0.9)</td>
</tr>
</tbody>
</table>

Specificity (95% CI)
Fig S3B

Specificity (95% CI)

- Caminero 1993: 1.00 (0.93 - 1.0)
- Kunter 2003 (a): 0.86 (0.71 - 0.9)
- Kunter 2003 (b): 0.92 (0.78 - 0.9)
- Kunter 2003 (c): 0.81 (0.65 - 0.9)

Specificity

0 0.2 0.4 0.6 0.8 1