Commercial Serological Antibody Detection Tests for the Diagnosis of Pulmonary Tuberculosis: A Systematic Review

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

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 County of Sacramento Department of Health and Human Services, Sacramento, California, United States of America, 4 Department of Pathology, 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 Veterans Affairs Medical Center, New York, United States of America, 7 UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases, World Health Organization, Geneva, Switzerland, 8 Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Canada, 9 Respiratory Epidemiology and Clinical Research Unit, Montréal Chest Institute, Montréal, Canada, 10 Statens Serum Institut, Department of Infectious Disease Immunology, Copenhagen, Denmark

Abbreviations: AUC, area under the curve; CI, confidence interval; ELISA, enzyme-linked immunosorbent assay; FPR, false positive rate; LAM, lipoarabinomannan; ROC, receiver operating characteristic; SROC, summary receiver operating characteristic; TB, tuberculosis; TPR, true positive rate

* To whom correspondence should be addressed. E-mail: madhukar. pai@mcgill.ca

ABSTRACT

Background

The global tuberculosis epidemic results in nearly 2 million deaths and 9 million new cases of the disease a year. The vast majority of tuberculosis patients live in developing countries, where the diagnosis of tuberculosis relies on the identification of acid-fast bacilli on unprocessed sputum smears using conventional light microscopy. Microscopy has high specificity in tuberculosis-endemic countries, but modest sensitivity which varies among laboratories (range 20% to 80%). Moreover, the sensitivity is poor for paucibacillary disease (e.g., pediatric and HIV-associated tuberculosis). Thus, the development of rapid and accurate new diagnostic tools is imperative. Immune-based tests are potentially suitable for use in low-income countries as some test formats can be performed at the point of care without laboratory equipment. Currently, dozens of distinct commercial antibody detection tests are sold in developing countries. The question is "do they work?"

Methods and Findings

We conducted a systematic review to assess the accuracy of commercial antibody detection tests for the diagnosis of pulmonary tuberculosis. Studies from all countries using culture and/ or microscopy smear for confirmation of pulmonary tuberculosis were eligible. Studies with fewer than 50 participants (25 patients and 25 control participants) were excluded. In a comprehensive search, we identified 68 studies. The results demonstrate that (1) overall, commercial tests vary widely in performance; (2) sensitivity is higher in smear-positive than smear-negative samples; (3) in studies of smear-positive patients, Anda-TB IgG by enzyme-linked immunosorbent assay shows limited sensitivity (range 63% to 85%) and inconsistent specificity (range 73% to 100%); (4) specificity is higher in healthy volunteers than in patients in whom tuberculosis disease is initially suspected and subsequently ruled out; and (5) there are insufficient data to determine the accuracy of most commercial tests in smear microscopy–negative patients, as well as their performance in children or persons with HIV infection.

Conclusions

None of the commercial tests evaluated perform well enough to replace sputum smear microscopy. Thus, these tests have little or no role in the diagnosis of pulmonary tuberculosis. Lack of methodological rigor in these studies was identified as a concern. It will be important to review the basic science literature evaluating serological tests for the diagnosis of pulmonary tuberculosis to determine whether useful antigens have been described but their potential has not been fully exploited. Activities leading to the discovery of new antigens with immunodiagnostic potential need to be intensified.

The Editors' Summary of this article follows the references.

Introduction

The burden of disability and death due to tuberculosis (TB) is immense, with 8.8 million new cases of the disease and 1.6 million deaths estimated to have occurred in 2005 alone [1]. Although the incidence of TB is constant or falling in many regions of the world, rates remain high in sub-Saharan Africa as a consequence of the HIV epidemic [1,2]. The expansion of DOTS, the centerpiece of the international TB partnership control strategy, has resulted in improved case-detection rates during the past several years; however, the majority of DOTS programs in high-burden countries have fallen short of the 2005 global target of 70% case detection of the most infectious cases [1].

The vast majority of TB patients live in low- and middle-income countries [2], where the diagnosis of TB disease relies primarily on identification of acid-fast bacilli on unprocessed sputum smears using a conventional light microscope. Microscopy is highly specific for *Mycobacterium tuberculosis* in TB-endemic countries [3,4]. Although microscopy has been reported to have greater than 80% sensitivity for identifying cases of pulmonary TB in some settings [5,6], the sensitivity of the test has been lower and variable in other reports (range 20% to 80%) [4,7]. Moreover, sensitivity is poor for paucibacillary disease (e.g., pediatric and HIV-associated TB) [8,9], a major concern on account of the strong association between HIV infection and smear negativity [10,11]. This lack of sensitivity of the sole diagnostic test in many parts of the world results in delays in diagnosis, enabling the disease to progress and increasing the potential for transmission of *M. tuberculosis* [5]. To ensure appropriate care for patients and to improve control of the global TB epidemic, simple, accurate, inexpensive, and, ideally, point-of-care diagnostic tools for TB are urgently needed.

The relative importance of the different characteristics of a diagnostic test depends upon the setting in which the test is to be performed and the intended use of the results. Technical simplicity, for example, is essential if a test is to be used in a primary health-care clinic or basic health laboratory in low-income countries. If test results are to be used to exclude a diagnosis of TB in patients with respiratory symptoms in TB-endemic countries, then tests with a high sensitivity (high negative predictive value) are required even if the test is only moderately specific. Excluding TB patients from this group would then allow a more rigorous diagnostic work-up to be performed on a smaller group of patients. On the other hand, if a test is to be used to identify patients with respiratory symptoms in endemic countries for anti-TB treatment, a high specificity (high positive predictive value) is required. In the latter case, high sensitivity would also be desirable.

Immune-based tests would seem to offer the potential to improve case detection as currently performed, as some of the test formats (e.g., immunochromatographic test) are suitable for resource-limited areas. The major advantages of immune-based tests are their speed (results may be available within hours) and simplicity compared with microscopy [8,12]. The development of immune-based tests for the detection of antibodies, antigens, and immune complexes has been attempted for decades, and their performance has been critically appraised in several descriptive reviews and textbook chapters [13–22]. The most common of these tests rely on detection of the humoral (serological) antibody immune response to *M. tuberculosis* (the subject of this systematic 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 [23,24] and pleural fluid [25]).

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, in developing countries, where diagnostic tests are rarely subjected to regulatory review or approval [26,27], test manufacturers and distributors are marketing dozens of different antibody detection diagnostic commercial kits. The extent of their use is largely unknown; however, companies report sales volumes between 3,000 and 300,000 tests per year [28]. These tests differ in a number of their features, including antigen composition, antigen source (e.g., native or recombinant), chemical composition (e.g., protein, carbo-hydrate, or lipid), extent and manner of purification of the antigen(s), and class of immunoglobulin detected (e.g., IgG, IgM, or IgA). Performance data are often limited to those found in the package inserts and, being based on small sets of patients, are typically favorable (J. Cunningham, personal communication).

An antibody detection test can be developed into a number of formats depending on the membrane, antigen(s) coating, and incubation technique. Common designs include the enzyme-linked immunosorbent assay (ELISA) format and the immunochromatographic test format. ELISA is a complex assay with several steps: coating of antigens onto the surface of plastic wells, the addition of serum samples to the wells, and several washing stages. Antigen–antibody reactions are visualized using anti-human antibody linked to an enzymatic indicator system [29]. The assay can take hours to perform. The immunochromatographic test is simpler. In this technique, the antigens are pre-coated in lines across a membrane (e.g., nitrocellulose) to which samples are applied. Antigen–antibody reactions are visualized on the lines using anti-human antibody bound to substances such as colloidal gold. The test takes only a few minutes to perform [30].

The ELISA format has the advantages that many serum samples can be tested in parallel and the process can be completely automated, making this technique attractive in fully equipped laboratories that test a large number of samples. For developing countries with limited laboratory resources and access, an immunochromatographic test would be the pre-ferred method, as this format requires only visual inspection of the antigen-containing lines and can, therefore, be performed at the point of care without laboratory equipment.

An initial survey of the literature found more than 200 studies that have evaluated commercial serological antibody detection tests, hereafter referred to as commercial tests, for the diagnosis of pulmonary TB. To our knowledge, this vast body of literature has not been systematically reviewed and 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 [31]. We specifically addressed two questions. (1) How accurate are commercial tests for the diagnosis of pulmonary TB overall and for smear-positive and smear-negative disease? (2) What is

the specificity of commercial tests in healthy control participants compared with the specificity in patients without TB, but in whom TB was initially suspected?

Methods

Search Strategy

We searched electronic databases for primary studies: PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed) (1990 to May 2006), BIOSIS (http://scientific.thomson.com/biosis/) (1990 to October 2005), Embase (http://www.embase.com) (1990 to October 2005), and Web of Science (http://scientific.thomson.com/products/wos/) (1990 to October 2005). The search terms used included the following: "tuberculosis", "*Mycobacterium tuberculosis*", "immunological tests", "serological tests", "antibody detection", "antigen detection", "ELISA", "Western blot", and "sensitivity and specificity". We identified additional studies by contacting experts in the field and by searching reference lists from primary studies, review articles, and textbook chapters.

Study Selection

Our search strategy sought to identify all available articles published in English that evaluated commercial tests for the serological diagnosis of pulmonary TB. We included only those studies in which patients had bacteriologically confirmed pulmonary TB, and in which results were provided separately for smear-positive and smear-negative patients. In particular, we defined the reference standard as either isolation of *M. tuberculosis* on culture, or, for studies conducted without culture in endemic countries, the presence of acid-fast bacilli detected by sputum smear microscopy. No restrictions were made with respect to study design (cross sectional or case control) or data collection (prospective or retrospective).

We excluded studies that relied solely on clinical or radiological features or improvement while on anti-TB therapy as the criteria for establishing the diagnosis of TB. In addition, the following studies were excluded: (1) studies published before 1990, for the reason that many studies used crude antigen extracts or obsolete immunological methods; (2) studies with fewer than 50 participants (at least 25 TB patients and 25 control participants were required for inclusion); (3) studies of latent TB infection; (4) studies of nontuberculous mycobacteria; (5) studies of antibody responses during or after TB treatment; (6) investigations conducted using non-immunologic methods for detection of antibodies; (7) basic science literature that focuses on cloning of new antigens or their immunologic properties (e.g., epitope mapping) or other new methods of antibody detection; and (8) case reports and reviews.

Initially, two reviewers (KRS and MH) screened citations retrieved from all sources. To identify relevant studies pertaining specifically to commercial tests, a second screening was done (KRS and MH) of full texts from citations found to be relevant in the first screen. A list of excluded studies, along with the reasons for exclusion, is available from the authors on request.

Data Extraction

We created and piloted a data-extraction form with a subset of eligible studies. Based upon experience gained in the pilot, the data-extraction form was finalized. One reviewer (KRS) extracted data from all eligible studies on the following qualities: study design and methodological quality (see assessment of study quality below), study population, antigen and antibody characteristics, laboratory technique, reference standard, and outcome data (sensitivity and specificity). To verify reproducibility of data extraction, a second reviewer (M. Henry) independently extracted data from 15% of the included studies. The inter-rater agreement between the two reviewers for sensitivity and specificity estimates was 100%. When data were not clearly reported, the information was 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 (control participants) for each study in the following order: (1) patients in whom pulmonary TB was initially suspected but who were later found to have nontuberculous respiratory disease; (2) patients diagnosed with a variety of diseases other than TB (mixed disease); (3) healthy persons from endemic countries; (4) contacts of patients with TB; (5) mixed groups from categories (1) to (4); and (6) healthy persons from non-endemic countries. In our view, this hierarchy gave priority to the populations expected to be encountered in a routine clinical setting.

Assessment of Study Quality

We assessed the quality of studies using the following criteria, which have been suggested as being important for diagnostic studies [31]. (1) Was there a comparison of the commercial test with an independent, appropriate reference standard (i.e., the commercial test did not form part of the reference standard)? (2) Was the commercial test result performed and recorded by technicians who were unaware (blinded) of the results of the reference standard? (3) Did the whole sample or a randomly selected subset of the sample receive verification using the reference standard? (4) Did the study prospectively recruit consecutive patients suspected of having pulmonary TB?

Data Collation and Meta-Analysis

We used standard methods recommended for meta-analyses of diagnostic test evaluations [31,32]. As studies were heterogeneous, particularly with respect to the antigen composition of the tests, antibody class (IgG, IgM, or IgA), comparison groups, and sputum status of the patients, we first grouped studies by type of commercial test and then further stratified by immunoglobulin class and smear status. To calculate sensitivity and specificity values for the commercial tests, we cross-tabulated each result against the reference standard. Sensitivity refers to the proportion of pulmonary TB patients with a positive result on a given commercial test; specificity refers to the proportion of TB-negative participants who had negative results on a given commercial test. Whenever possible, we extracted raw data from primary studies to fill the four cell values of a diagnostic 2×2 table: true positives, false positives, true negatives, and false negatives. In calculations of sensitivity, we included studies from endemic countries that used sputum smear positivity as the reference standard along with studies using culture as the reference standard. We recognized that some
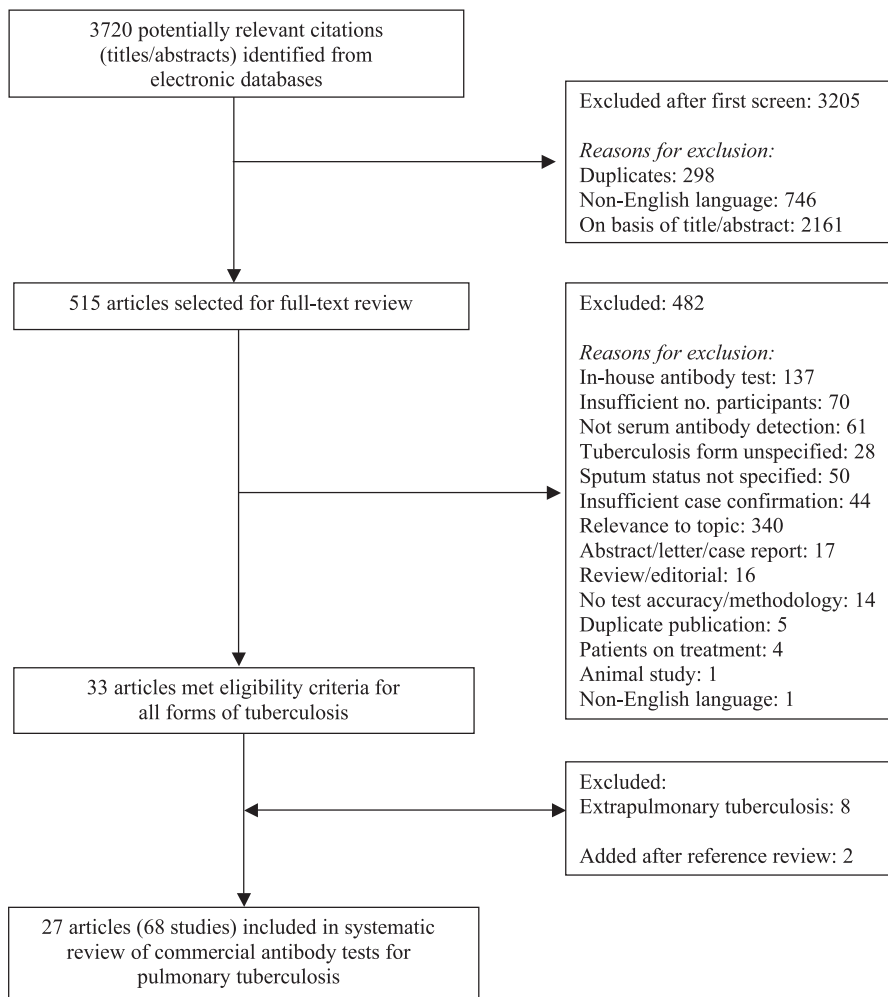
Figure 1. Flow Diagram for Study Selection
doi:10.1371/journal.pmed.0040202.g001

authors used the same comparison group for multiple studies, and thus derived identical specificity estimates. Therefore, in determining specificity, we included the specificity value for the specific comparison group only once when appropriate. For clarity of presentation, studies that reported results stratified by subgroups are shown more than once in tables or figures.

Data were analyzed using SPSS (version 14.0.1.366) [33] and Meta-DiSc (version 1.4) software [34]. Sensitivity and specificity values were calculated for the commercial tests investigated in each study, along with their 95% confidence intervals (CIs). In addition to the sensitivity and specificity estimates and forest plots generated for this review, true positive rates (TPR = sensitivity) and false positive rates (FPR = 1 − specificity) were summarized using an asymmetric summary receiver operating characteristic (SROC) curve [35]. TPR and FPR are not independent of each other as they vary with the thresholds (cut points for determining test positives) employed in the original studies. In addition, it is likely that different thresholds were used in various studies, either implicitly or explicitly. Because of the inherent trade-off between TPR and FPR, it is imperative to plot the estimates of the two quantities in a receiver operating characteristic

(ROC) space and to use meta-analytic methods that take into account the threshold effect. Thus, we did not pool the sensitivity and specificity estimates separately; instead we analyzed TPR and FPR as pairs in an SROC analysis, and explored the effect of variability in cut points on study results.

Unlike a traditional ROC plot that explores the effect of varying thresholds on sensitivity and specificity in a single study, each data point in the SROC space represents an individual study. As described by Littenberg and Moses [32], the SROC analysis involves three steps: (1) the pairs of TPR and FPR estimates from each study are transformed onto a suitable scale of log odds; (2) a linear regression equation is fitted using the transformed data; and (3) the coefficients from the linear regression model are used to generate a curve in the original ROC space.

The area under the curve (AUC) (in this case, being the area under the SROC curve) presents an overall summary of test performance and displays the trade-off between sensitivity and specificity. An AUC of 1.0 (100%) indicates perfect discriminatory ability of the diagnostic test. In addition, the Q* index is another useful global estimate of test accuracy for comparing SROC curves. The Q* index, defined by the point

Table 1. Commercial Antibody Detection Tests for the Diagnosis of Pulmonary TB

Name of Test (Number of Studies in the Review)[a]	Antigen(s)	Antigen Source	Immuno-globulin Class	Laboratory Technique	Name of Manufacturer	Address (URL)
Anda-TB (19)	Antigen 60	Native	IgG, IgA, IgM	ELISA	Anda Biologicals	Strasbourg, France (http://www.andabiologicals.com)
Detect-TB (2)	Five different proteins	Three recombinant antigens and two synthetic peptides secreted by *M. tuberculosis* H37Rv	IgG	ELISA	BioChem ImmunoSystems	Montréal, Quebec, Canada
ICT TB Test (3)	38 kDa and four proprietary antigens	Recombinant	IgG	Immunochromatographic test	ICT Diagnostics	Balgowlah, New South Wales, Australia
Kaolin agglutination test (2)	Tuberculophosphatide	Native	IgG	Kaolin agglutination test	Hitech Laboratories	Bombay, India
MycoDot (5)	LAM	Native	IgG	Immunodot rapid (20 min) test	Mossman Associates	Blackstone, Massachusetts, USA (http://www.mossmanassociates.com)
Pathozyme Myco (12)	LAM and 38 kDa	Native LAM and recombinant 38 kDa	IgG, IgM, and IgA	ELISA	Omega Diagnostics	Alloa, Scotland, UK (http://www.omegadiagnostics.co.uk)
Pathozyme TB Complex Plus (5)	38 kDa and 16 kDa	Recombinant 38 kDa and recombinant 16 kDa	IgG	ELISA	Omega Diagnostics	Alloa, Scotland, UK (http://www.omegadiagnostics.co.uk)
TB enzyme immunoassay (5)	Kp-90 antigenic compound	Native sonicated preparation	IgG and IgA	ELISA	Kreatech	Amsterdam, The Netherlands
TB glycolipid assay (6)	Contains: trehalose 6,6'-dimycolate and trehalose monomycolate, diacyltrehalose, phenolic glycolipid, 2,3,6,6-tetraacyl-trehalose-2-sulfate, and 2,3,6-triacyl-trehalose	Native preparation from the cell walls of *M. tuberculosis* H37Rv	IgG	ELISA	Kyowa Medex	Tokyo, Japan, (http://www.kyowamx.co.jp)

Some of the commercial tests listed may no longer be manufactured and distributed.

[a]Anda-TB tests include: IgG (14); IgM (2); IgA (2); IgG and IgM (1). Pathozyme Myco tests include: Myco G (3); Myco M (3); Myco A (2); Myco G and M (1); Myco G and A (1); Myco M and A (1); Myco G, M, and A (1). TB enzyme immunoassay includes: IgA (4); IgG (1). Nine studies used various combinations of Pathozyme Myco and Pathozyme TB Complex Plus.

doi:10.1371/journal.pmed.0040202.t001

Table 2. Selected Characteristics of Studies Investigating Anda-TB for the Diagnosis of Pulmonary TB

Author (Year, Study)	Data Collection	Patient Selection	Verification	Reference Standard/ Smear Status	Patient Country	Type of Comparison Group	Immuno-globulin Class	Number of Participants[b]	Sensitivity (95% CI)	Specificity (95% CI)
Al-Hajjaj (1999, a) [38]	Retrospective	Consecutive	Complete	Smear/smear positive	Saudi Arabia	Nontuberculous respiratory	IgG	200/106	0.77 (0.70–0.82)	0.92 (0.86–0.97)
Al-Hajjaj (1999, b) [38]	Retrospective	Consecutive	Complete	Smear/smear positive	Saudi Arabia	Nontuberculous respiratory	IgM	200/106	0.51 (0.43–0.58)	1.00 (0.97–1.00)
Al-Hajjaj (1999, c) [38]	Retrospective	Consecutive	Complete	Smear/smear positive	Saudi Arabia	Nontuberculous respiratory and healthy	IgG/IgM	200/181	0.87 (0.81–0.91)	0.95 (0.91–0.98)
Alifano (1994) [40]	Retrospective	NR	Differential	Culture/smear positive	Italy	Healthy	IgG	42/94	0.83 (0.69–0.93)	0.98 (0.93–1.00)
Alifano (1996, a) [41]	Retrospective	NR	NR	Culture/smear positive	Italy	Nontuberculous respiratory	IgG	33/44	0.85 (0.68–0.95)	0.93 (0.81–0.99)
Alifano (1996, b) [41]	Retrospective	NR	NR	Culture/smear positive	Italy	Nontuberculous respiratory	IgA	33/44	0.82 (0.65–0.93)	0.89 (0.75–0.96)
Alifano (1996, c) [41]	Retrospective	NR	NR	Culture/smear negative	Italy	Nontuberculous respiratory	IgG	36/44	0.64 (0.46–0.79)	0.93 (0.81–0.99)
Alifano (1996, d) [41]	Retrospective	NR	NR	Culture/smear negative	Italy	Nontuberculous respiratory	IgA	36/44	0.64 (0.46–0.79)	0.89 (0.75–0.96)
Kalantri (2005, a) [51][a]	Retrospective	NR	Complete	Culture/smear positive	India	Nontuberculous respiratory	IgG	105/40	0.80 (0.71–0.87)	1.00 (0.91–1.00)
Kalantri (2005, b) [51][a]	Retrospective	NR	Complete	Culture/smear positive	India	Nontuberculous respiratory	IgM	105/40	0.29 (0.20–0.38)	1.00 (0.91–1.00)
Luh (1996, a) [52]	Prospective	NR	NR	Smear/smear positive	Taiwan	Nontuberculous respiratory	IgG	62/293	0.84 (0.72–0.92)	0.89 (0.85–0.92)
Luh (1996, b) [52]	Prospective	NR	NR	Culture/smear negative	Taiwan	Nontuberculous respiratory	IgG	70/293	0.71 (0.59–0.82)	0.89 (0.85–0.92)
Okuda (2004, a) [56][a]	Prospective	NR	Differential	Culture/smear positive	Japan	Nontuberculous respiratory	IgG	34/111	0.82 (0.65–0.93)	0.91 (0.84–0.96)
Okuda (2004, b) [56][a]	Prospective	NR	Differential	Culture/smear negative	Japan	Nontuberculous respiratory	IgG	26/111	0.73 (0.52–0.88)	0.91 (0.84–0.96)
Sachan (1994) [59]	Retrospective	Consecutive	Differential	Smear/smear positive	India	Contacts/healthy	IgG	66/32	0.85 (0.74–0.92)	1.00 (0.89–1.00)
Traumuller (2005) [61]	Retrospective	NR	Complete	Culture/smear positive	Saudi Arabia	Nontuberculous respiratory	IgG	38/79	0.84 (0.69–0.94)	0.73 (0.62–0.83)
van der Werf (1992) [62]	Retrospective	Consecutive	NR	Smear/smear positive	Ghana	Mixed disease	IgG	46/30	0.78 (0.64–0.89)	0.87 (0.69–0.96)
Wu, HP (2004, a) [64]	Retrospective	NR	Complete	Culture/smear positive	Taiwan	Nontuberculous respiratory	IgG	92/34	0.63 (0.52–0.73)	0.88 (0.73–0.97)
Wu, HP (2004, b) [64]	Retrospective	NR	Complete	Culture/smear negative	Taiwan	Nontuberculous respiratory	IgG	86/34	0.35 (0.25–0.46)	0.88 (0.73–0.97)

Nineteen studies.
Anda-TB (Anda Biologicals, Strasbourg, France).
[a]Blinded studies.
[b]Number of participants with TB/without TB.
NR, not reported.
doi:10.1371/journal.pmed.0040202.t002

Table 3. Selected Characteristics of Studies Investigating TB Enzyme Immunoassay for the Diagnosis of Pulmonary TB

Author (Year, Study)	Study Direction	Patient Selection	Verification	Reference Standard/ Smear Status	Patient Country	Type of Comparison Group	Immuno-globulin Class	Number of Participants[b]	Sensitivity (95% CI)	Specificity (95% CI)
Alifano (1997, a) [39]	Retrospective	Consecutive	Differential	Culture/smear positive	Italy	Nontuberculous respiratory	IgA	32/28	0.75 (0.57–0.89)	0.86 (0.67–0.96)
Alifano (1997, b) [39]	Retrospective	Consecutive	Differential	Culture/smear negative	Italy	Nontuberculous respiratory	IgA	56/28	0.68 (0.54–0.80)	0.86 (0.67–0.96)
Conde (2004, a) [46][a]	Prospective	NR	Complete	Culture/smear positive	Brazil	Nontuberculous respiratory	IgG	40/31	0.75 (0.59–0.87)	0.77 (0.59–0.90)
Conde (2004, b) [46][a]	Prospective	NR	Complete	Culture/smear positive	Brazil	Nontuberculous respiratory	IgA	40/31	0.83 (0.67–0.93)	0.48 (0.30–0.67)
Julián (2000) [50]	Prospective	NR	Complete	Culture/mear positive	Spain	Nontuberculous respiratory	IgA	27/36	0.74 (0.54–0.89)	0.47 (0.30–0.65)

Five studies.
TB Enzyme Immunoassay (Kreatech Diagnostics, Amsterdam, The Netherlands).
[a]Blinded studies.
[b]Number of participants with TB/without TB.
NR, not reported.
doi:10.1371/journal.pmed.0040202.t003

Table 4. Selected Characteristics of Studies Investigating Pathozyme TB Complex Plus for the Diagnosis of Pulmonary TB

Author (Year, Study)	Study Direction	Patient Selection	Verification	Reference Standard/ Smear Status	Patient Country	Type of Comparison Group	Number of Participants[b]	Sensitivity (95% CI)	Specificity (95% CI)
Butt (2004, c) [44]	Retrospective	NR	Complete	Smear/smear positive	Pakistan	Healthy	94/90	0.64 (0.53–0.73)	0.97 (0.91–0.99)
Imaz (2004, h) [48][a]	Prospective	Consecutive	Complete	Culture/smear negative	Argentina	Mixed disease	41/45	0.29 (0.16–0.46)	1.00 (0.92–1.00)
Julián (2004, d) [49]	Retrospective	NR	Differential	Culture/smear positive	Spain	Nontuberculous respiratory	29/35	0.31 (0.15–0.51)	0.86 (0.70–0.95)
Wilkinson (1997, a) [63]	Retrospective	NR	NR	Culture/smear positive	India	Healthy	125/34	0.78 (0.70–0.85)	0.94 (0.80–0.99)
Wilkinson (1997, b) [63]	Retrospective	NR	NR	Culture/smear negative	India	Healthy	71/34	0.76 (0.64–0.85)	0.94 (0.80–0.99)

Five studies.
Pathozyme TB Complex Plus (Omega Diagnostics, Alloa, Scotland, UK) detects immunoglobulin G.
[a]Blinded studies.
[b]Number of participants with TB/without TB.
NR, not reported.
doi:10.1371/journal.pmed.0040202.t004

Table 5. Selected Characteristics of Studies Investigating TB Glycolipid Assay for the Diagnosis of Pulmonary TB

Author (Year, Study)	Study Direction	Patient Selection	Verification	Reference Standard/ Smear Status	Patient Country	Type of Comparison Group	Number of Participants[b]	Sensitivity (95% CI)	Specificity (95% CI)
Iinuma (2002) [47]	Retrospective	NR	Differential	Smear/smear positive	Japan	Nontuberculous respiratory	47/54	0.89 (0.77–0.96)	0.72 (0.58–0.84)
Maekura (2001, a) [54][a]	Retrospective	NR	Differential	Culture/smear positive	Japan	Nontuberculous respiratory	164/180	0.90 (0.84–0.94)	0.85 (0.79–0.90)
Maekura (2001, b) [54][a]	Retrospective	NR	Differential	Culture/smear negative	Japan	Nontuberculous respiratory	52/180	0.69 (0.55–0.81)	0.85 (0.79–0.90)
Maekura (2003) [53][a]	Prospective	Consecutive	Complete	Culture/smear negative	Japan	Nontuberculous respiratory	70/52	0.60 (0.48–0.72)	0.88 (0.77–0.96)
Okuda (2004, e) [56][a]	Prospective	NR	Differential	Culture/smear positive	Japan	Nontuberculous respiratory	34/111	0.76 (0.59–0.89)	0.89 (0.82–0.94)
Okuda (2004, f) [56][a]	Prospective	NR	Differential	Culture/smear negative	Japan	Nontuberculous respiratory	26/111	0.77 (0.56–0.91)	0.89 (0.82–0.94)

Six studies.
TB Glycolipid Assay (Kyowa Medex, Tokyo, Japan) detects immunoglobulin G.
[a]Blinded studies.
[b]Number of participants with TB/without TB.
NR, not reported.
doi:10.1371/journal.pmed.0040202.t005

Table 6. Selected Characteristics of Studies Investigating ICT for the Diagnosis of Pulmonary TB

Author (Year, Study)	Study Direction	Patient Selection	Verification	Reference Standard/ Smear Status	Patient Country	Type of Comparison Group	Number of Participants[b]	Sensitivity (95% CI)	Specificity (95% CI)
McConkey (2002) [55][a]	Prospective	NR	Complete	Culture/smear positive	Egypt	Nontuberculous respiratory	71/74	0.87 (0.77–0.94)	0.82 (0.72–0.90)
Ongut (2006) [57]	Retrospective	NR	Differential	Culture/smear positive	Turkey	Healthy	53/54	0.40 (0.26–0.54)	1.00 (0.93–1.00)
Perkins (2003) [58]	Retrospective	NR	Complete	Smear/smear positive	Brazil	Nontuberculous respiratory	120/34	0.64 (0.55–0.73)	0.85 (0.69–0.95)

Three studies.
ICT (ICT Diagnostics, Balgowlah, New South Wales, Australia) detects immunoglobulin G.
[a]Blinded studies.
[b]Number of participants with TB/without TB.
NR, not reported.
doi:10.1371/journal.pmed.0040202.t006

Table 7. Selected Characteristics of Studies Investigating MycoDot for the Diagnosis of Pulmonary TB

Author (Year, Study)	Study Direction	Patient Selection	Verification	Reference Standard/ Smear Status	Patient Country	Type of Comparison Group	Number of Participants[b]	Sensitivity (95% CI)	Specificity (95% CI)
Antunes (2002, a) [43][a]	Prospective	Random	Complete	Smear/smear positive	Guinea-Bissau	Mixed disease	46/223	0.63 (0.48–0.77)	0.92 (0.88–0.95)
Antunes (2002, b) [43][a,c]	Prospective	Random	Complete	Smear/smear positive	Guinea-Bissau	Mixed disease	34/198	0.68 (0.49–0.83)	0.92 (0.87–0.95)
Okuda (2004, c) [56][a]	Prospective	NR	Differential	Culture/smear positive	Japan	Nontuberculous respiratory	34/111	0.76 (0.59–0.89)	0.97 (0.92–0.99)
Okuda (2004, d) [56][a]	Prospective	NR	Differential	Culture/smear negative	Japan	Nontuberculous respiratory	26/111	0.58 (0.37–0.77)	0.97 (0.92–0.99)
Somi (1999) [60][a,d]	Prospective	NR	Complete	Culture/smear positive	Tanzania	Nontuberculous respiratory	39/102	0.26 (0.13–0.42)	0.84 (0.76–0.91)

Five studies.
MycoDot (Mossman Associates, Blackstone, Massachusetts, USA) detects immunoglobulin G.
[a]Blinded studies.
[b]Number of participants with TB/without TB.
[c]HIV-negative individuals only, substudy of Antunes study (a) which includes both HIV-positive and HIV-negative individuals.
[d]Commercial test performed on whole blood.
NR, not reported.
doi:10.1371/journal.pmed.0040202.t007

Table 8. Selected Characteristics of Studies Investigating Pathozyme Myco for the Diagnosis of Pulmonary TB

Author (Year, Study)	Study Direction	Patient Selection	Verification	Reference Standard/Smear Status	Patient Country	Type of Comparison Group	Immunoglobulin Class	Number of Participants[b]	Sensitivity (95% CI)	Specificity (95% CI)
Butt (2004, a) [44]	Retrospective	NR	Complete	Smear/smear positive	Pakistan	Healthy	IgG	94/90	0.46 (0.35–0.56)	0.93 (0.86–0.98)
Butt (2004, b) [44]	Retrospective	NR	Complete	Smear/smear positive	Pakistan	Healthy	IgM	94/90	0.67 (0.57–0.76)	0.98 (0.92–1.00)
Imaz (2004, a) [48][a]	Prospective	Consecutive	Complete	Culture/smear negative	Argentina	Mixed disease	IgG	41/45	0.49 (0.33–0.65)	1.00 (0.92–1.00)
Imaz (2004, b) [48][a]	Prospective	Consecutive	Complete	Culture/smear negative	Argentina	Mixed disease	IgM	41/45	0.32 (0.18–0.48)	0.93 (0.82–0.99)
Imaz (2004, c) [48][a]	Prospective	Consecutive	Complete	Culture/smear negative	Argentina	Mixed disease	IgA	41/45	0.34 (0.20–0.51)	0.98 (0.88–1.00)
Imaz (2004, d) [48][a]	Prospective	Consecutive	Complete	Culture/smear negative	Argentina	Mixed disease	IgG and IgM	41/45	0.63 (0.47–0.78)	0.93 (0.82–0.99)
Imaz (2004, e) [48][a]	Prospective	Consecutive	Complete	Culture/smear negative	Argentina	Mixed disease	IgG and IgA	41/45	0.59 (0.42–0.74)	0.98 (0.88–1.00)
Imaz (2004, f) [48][a]	Prospective	Consecutive	Complete	Culture/smear negative	Argentina	Mixed disease	IgM and IgA	41/45	0.49 (0.33–0.65)	0.91 (0.79–0.98)
Imaz (2004, g) [48][a]	Prospective	Consecutive	Complete	Culture/smear negative	Argentina	Mixed disease	IgG, IgM, and IgA	41/45	0.68 (0.52–0.82)	0.91 (0.79–0.98)
Julián (2004, a) [49]	Retrospective	NR	Differential	Culture/smear positive	Spain	Nontuberculous respiratory	IgG	29/35	0.41 (0.24–0.61)	0.97 (0.85–1.00)
Julián (2004, b) [49]	Retrospective	NR	Differential	Culture/smear positive	Spain	Nontuberculous respiratory	IgM	29/35	0.10 (0.02–0.27)	0.97 (0.85–1.00)
Julián (2004, c) [49]	Retrospective	NR	Differential	Culture/smear positive	Spain	Nontuberculous respiratory	IgA	29/35	0.21 (0.08–0.40)	0.97 (0.85–1.00)

Twelve studies.
Pathozyme Myco (Omega Diagnostics, Alloa, Scotland, UK).
[a]Blinded studies.
[b]Number of participants with TB/without TB.
NR, not reported.
doi:10.1371/journal.pmed.0040202.t008

Table 9. Selected Characteristics of Studies Investigating Pathozyme Myco and Pathozyme TB Complex Plus in Various Combinations for the Diagnosis of Pulmonary TB

Author (Year, Study)	Study Direction	Patient Selection	Verification	Reference Standard/Smear Status	Patient Country	Type of Comparison Group	Antibody Test[b] Combination	Number of Participants[c]	Sensitivity (95% CI)	Specificity (95% CI)
Butt (2004, d) [44]	Retrospective	NR	Complete	Smear/smear positive	Pakistan	Healthy	Pathozyme Myco G and Pathozyme TB Complex Plus	94/90	0.76 (0.66–0.84)	0.93 (0.86–0.98)
Butt (2004, e) [44]	Retrospective	NR	Complete	Smear/smear positive	Pakistan	Healthy	Pathozyme Myco M and Pathozyme TB Complex Plus	94/90	0.82 (0.73–0.89)	0.97 (0.91–0.99)
Butt (2004, f) [44]	Retrospective	NR	Complete	Smear/smear positive	Pakistan	Healthy	Pathozyme Myco G, Myco M, and Pathozyme TB Complex Plus	94/90	0.87 (0.79–0.93)	0.93 (0.86–0.98)
Imaz (2004, i) [48][a]	Prospective	Consecutive	Complete	Culture/smear negative	Argentina	Mixed disease	Pathozyme Myco G and Pathozyme TB Complex Plus	41/45	0.59 (0.42–0.74)	1.00 (0.92–1.00)
Imaz (2004, j) [48][a]	Prospective	Consecutive	Complete	Culture/smear negative	Argentina	Mixed disease	Pathozyme Myco M and Pathozyme TB Complex Plus	41/45	0.51 (0.35–0.67)	0.93 (0.82–0.99)
Imaz (2004, k) [48][a]	Prospective	Consecutive	Complete	Culture/smear negative	Argentina	Mixed disease	Pathozyme Myco A and Pathozyme TB Complex Plus	41/45	0.49 (0.33–0.65)	0.98 (0.88–1.00)
Imaz (2004, l) [48][a]	Prospective	Consecutive	Complete	Culture/smear negative	Argentina	Mixed disease	Pathozyme Myco G, Myco A, and Pathozyme TB Complex Plus	41/45	0.66 (0.49–0.80)	0.98 (0.88–1.00)
Imaz (2004, m) [48][a]	Prospective	Consecutive	Complete	Culture/smear negative	Argentina	Mixed disease	Pathozyme Myco A, Myco M, and Pathozyme TB Complex Plus	41/45	0.61 (0.45–0.76)	0.91 (0.79–0.98)
Imaz (2004, n) [48][a]	Prospective	Consecutive	Complete	Culture/smear negative	Argentina	Mixed disease	Pathozyme Myco G, Myco A, Myco M, and Pathozyme TB Complex Plus	41/45	0.76 (0.60–0.88)	0.91 (0.79–0.98)

Nine studies.
Pathozyme Myco and Pathozyme TB Complex Plus (both from Omega Diagnostics, Alloa, Scotland, UK).
Pathozyme TB Complex Plus detects IgG antibody.
[a]Blinded studies.
[b]Myco G, M, and A detect IgG, IgM, and IgA antibodies, respectively.
[c]Number of participants with TB/without TB.
NR, not reported.
doi:10.1371/journal.pmed.0040202.t009

Table 10. Selected Characteristics of Studies Investigating Detect-TB and Kaolin Agglutination Test for the Diagnosis of Pulmonary TB

Author (Year, Study)	Study Direction	Patient Selection	Verification	Reference Standard/ Smear Status	Patient Country	Type of Comparison Group	AntibodyTest[a]	Number of Participants[b]	Sensitivity (95% CI)	Specificity (95% CI)
Amicosante (1999, a) [42]	Retrospective	NR	NR	Culture/smear positive	Italy	Healthy	Detect-TB	54/150	0.83 (0.71–0.92)	0.97 (0.92–0.99)
Amicosante (1999, b) [42]	Retrospective	NR	NR	Culture/smear negative	Italy	Healthy	Detect-TB	46/150	0.65 (0.50–0.79)	0.97 (0.92–0.99)
Chandrase karan (1990, a) [45]	Prospective	NR	Complete	Smear/smear positive	India	Healthy	Kaolin agglutination test	184/208	0.54 (0.46–0.61)	0.86 (0.80–0.90)
Chandrase karan (1990, b) [45]	Prospective	NR	Complete	Culture/smear negative	India	Healthy	Kaolin agglutination test	66/208	0.55 (0.42–0.67)	0.86 (0.80–0.90)

Four studies.

Detect-TB (BioChem ImmunoSystems, Montréal, Canada).

Kaolin Agglutination Test (Hitech Laboratories, Bombay, India).

[a]Antibody tests detect immunoglobulin G.

[b]Number of participants with TB/without TB.

NR, not reported.

doi:10.1371/journal.pmed.0040202.t010

where sensitivity equals specificity on the SROC curve, is the point on the SROC curve that is intersected by the anti-diagonal. A Q* value of 1.0 indicates 100% accuracy (sensitivity and specificity of 100%) [32,35,36].

In meta-analyses of studies of diagnostic tests, heterogeneity refers to a high level of variability in study results [37]. Such heterogeneity could be a result of variability in thresholds, laboratory technique, disease spectrum, study design, and/or quality between studies [37]. In the presence of significant heterogeneity, pooled or summary estimates from meta-analyses are hard to interpret. Given the anticipated variability in accuracy, we decided, a priori, to avoid direct pooling of sensitivity and specificity values. Also, as described earlier, we addressed heterogeneity by using subgroup (stratified) analyses.

Results

Description of Included Studies

From the literature searches, we identified 3,720 citations, of which 27 publications (68 studies) met our eligibility criteria (Figure 1) [38–64]. We considered most studies to be distinct ([43], study b, is a substudy). Therefore, no effort was made to account for lack of independence. Of the total 68 studies, 32 (47%) collected data prospectively and 36 (53%), retrospectively. Twenty-four (35%) studies used either random or consecutive recruitment of participants, while 44 (65%) studies did not report the method of participant selection. Thirty-one (46%) studies reported at least single-blinded interpretation of commercial test and reference standard results. Mycobacterial culture was used as the gold standard in 51 (75%) studies and sputum smear microscopy was used in 17 (25%) studies. For all studies, the commercial test did not form part of the reference standard. In 39 (57%) studies, the entire study population was investigated using the identical reference standard (complete verification), while in 18 (26%) studies, the reference standard for TB patients and control participants differed (e.g., mycobacterial culture for TB patients and chest radiograph for control participants [differential verification]). For 11 (16%) studies, information about verification was unclear or not reported. Seventeen (25%) studies met all four criteria for good quality [43,48,53]. Forty-one (60%) studies were performed with smear microscopy–positive patients, and 27 (40%) were conducted with smear microscopy–negative patients. No studies involved children younger than 15 y old or patients with HIV infection. The median number of TB patients was 41 (interquartile range 38 to 75); the median number of control participants in comparison groups was 45 (interquartile range 40 to 107).

Antibody detection was done with stored, frozen sera in 59 (87%) studies and with fresh sera in one (2%) study [60]; in eight (12%) studies, the condition of the specimens was not reported. A total of nine different commercial tests are included in the review. For seven (78%) commercial tests, the specific antigen composition was described; for two (22%) tests, the antigen composition was considered a proprietary product. Fifty-nine (87%) studies assessed the performance of individual commercial tests. Nine (13%) studies evaluated the performance of two or more commercial tests used in combination. In these nine studies, the commercial tests (Pathozyme TB Complex Plus and Pathozyme Myco) came

	Sensitivity (95% CI)
Al-Hajjaj (1999) a	0.77 (0.70 - 0.82)
Al-Hajjaj (1999) b	0.51 (0.43 - 0.58)
Al-Hajjaj (1999) c	0.87 (0.81 - 0.91)
Alifano (1994)	0.83 (0.69 - 0.93)
Alifano (1996) a	0.85 (0.68 - 0.95)
Alifano (1996) b	0.82 (0.65 - 0.93)
Alifano (1996) c	0.64 (0.46 - 0.79)
Alifano (1996) d	0.64 (0.46 - 0.79)
Alifano (1997) a	0.75 (0.57 - 0.89)
Alifano (1997) b	0.68 (0.54 - 0.80)
Amicosante (1999) a	0.83 (0.71 - 0.92)
Amicosante (1999) b	0.65 (0.50 - 0.79)
Antunes (2002) a	0.63 (0.48 - 0.77)
Antunes (2002) b	0.68 (0.49 - 0.83)
Butt (2004) a	0.46 (0.35 - 0.56)
Butt (2004) b	0.67 (0.57 - 0.76)
Butt (2004) c	0.64 (0.53 - 0.73)
Butt (2004) d	0.76 (0.66 - 0.84)
Butt (2004) e	0.82 (0.73 - 0.89)
Butt (2004) f	0.87 (0.79 - 0.93)
Chandrasekaran (1990) a	0.54 (0.46 - 0.61)
Chandrasekaran (1990) b	0.55 (0.42 - 0.67)
Conde (2004) a	0.75 (0.59 - 0.87)
Conde (2004) b	0.83 (0.67 - 0.93)
Iinuma (2002)	0.89 (0.77 - 0.96)
Imaz (2004) a	0.49 (0.33 - 0.65)
Imaz (2004) b	0.32 (0.18 - 0.48)
Imaz (2004) c	0.34 (0.20 - 0.51)
Imaz (2004) d	0.63 (0.47 - 0.78)
Imaz (2004) e	0.59 (0.42 - 0.74)
Imaz (2004) f	0.49 (0.33 - 0.65)
Imaz (2004) g	0.68 (0.52 - 0.82)
Imaz (2004) h	0.29 (0.16 - 0.46)
Imaz (2004) i	0.59 (0.42 - 0.74)
Imaz (2004) j	0.51 (0.35 - 0.67)
Imaz (2004) k	0.49 (0.33 - 0.65)
Imaz (2004) l	0.66 (0.49 - 0.80)
Imaz (2004) m	0.61 (0.45 - 0.76)
Imaz (2004) n	0.76 (0.60 - 0.88)
Julian (2000)	0.74 (0.54 - 0.89)
Julian (2004) a	0.41 (0.24 - 0.61)
Julian (2004) b	0.10 (0.02 - 0.27)
Julian (2004) c	0.21 (0.08 - 0.40)
Julian (2004) d	0.31 (0.15 - 0.51)
Kalantri (2005) a	0.80 (0.71 - 0.87)
Kalantri (2005) b	0.29 (0.20 - 0.38)
Luh (1996) a	0.84 (0.72 - 0.92)
Luh (1996) b	0.71 (0.59 - 0.82)
Maekura (2001) a	0.90 (0.84 - 0.94)
Maekura (2001) b	0.69 (0.55 - 0.81)
Maekura (2003)	0.60 (0.48 - 0.72)
McConkey (2002)	0.87 (0.77 - 0.94)
Okuda (2004) a	0.82 (0.65 - 0.93)
Okuda (2004) b	0.73 (0.52 - 0.88)
Okuda (2004) c	0.76 (0.59 - 0.89)
Okuda (2004) d	0.58 (0.37 - 0.77)
Okuda (2004) e	0.76 (0.59 - 0.89)
Okuda (2004) f	0.77 (0.56 - 0.91)
Ongut (2006)	0.40 (0.26 - 0.54)
Perkins (2003)	0.64 (0.55 - 0.73)
Sachan (1994)	0.85 (0.74 - 0.92)
Somi (1999)	0.26 (0.13 - 0.42)
Traunmuller (2005)	0.84 (0.69 - 0.94)
van der Werf (1992)	0.78 (0.64 - 0.89)
Wilkinson (1997) a	0.78 (0.70 - 0.85)
Wilkinson (1997) b	0.76 (0.64 - 0.85)
Wu (2004) a	0.63 (0.52 - 0.73)
Wu (2004) b	0.35 (0.25 - 0.46)

0   0.2   0.4   0.6   0.8   1
Sensitivity

Figure 2. Sensitivity Estimates of Commercial Tests for the Diagnosis of Pulmonary TB
The circles and lines represent the point estimates and 95% CIs, respectively. The size of the circle indicates the study size.
doi:10.1371/journal.pmed.0040202.g002

from the same manufacturer, but differed in their antigen composition and/or immunoglobulin class. Table 1 lists selected characteristics of the commercial tests employed in the review, together with the names and addresses of their respective manufacturers.

How Accurate are Commercial Tests for the Diagnosis of Pulmonary TB Overall and for Smear-Positive and Smear-Negative Disease?

Tables 2–10 show performance and other selected characteristics for the commercial tests in the review. When all 68 studies were considered together, the sensitivity estimates ranged from 10% to 90% (Figure 2), and the specificity estimates ranged from 47% to 100% (Figure 3). On the whole, both measures varied widely among studies of a given commercial test and in studies across different commercial tests. Figure 4 shows performance for all commercial tests combined in a SROC curve. The AUC was 0.89 (95% CI 0.86–0.92), and the Q* index was 0.82, indicating modest accuracy.

The accuracy of commercial tests in patients with smear-positive pulmonary TB was higher (AUC = 0.90, 95% CI 0.86–0.94; 41 studies [Figure 5A]) than in patients with smear-negative disease (AUC = 0.84, 95% CI 0.77–0.91; 27 studies [Figure 5B]). These differences were, however, not statistically significant, based on overlapping CIs for the AUCs. Corresponding forest plots of sensitivity and specificity are shown in Figures S1–S4.

Anda-TB IgG was the test most frequently evaluated in the review. Figure 6 shows the forest plots for studies using Anda-TB IgG for the diagnosis of pulmonary TB in smear microscopy–positive patients (ten studies) [38,40,41,51,52,56,59,61,62,64]. Figure S5 shows the corresponding SROC curve. Sensitivity values ranged from 63% to 85%, with sensitivity ≤ 80% in four (40%) studies [38,51,62,64]; specificity values ranged from 73% to 100%, with specificity < 90% in four (40%) studies [52,61,62,64]. The AUC was 0.86 (95% CI 0.77–0.95), which was lower than the value noted in studies of smear microscopy–positive patients for all commercial tests combined. In smear microscopy–negative patients (four studies) [41,52,56,64], sensitivity estimates for Anda-TB IgG were low and variable (64%, 71%, 73%, and 35%); corresponding specificity estimates were similar (93%, 89%, 91%, and 88%) (Table 2). After stratifying by smear status, there were fewer than four studies for the other commercial tests. For each of these studies, sensitivity and specificity values are shown in Tables 2–10.

What is the Specificity of Commercial Tests in Healthy Control Participants Compared with the Specificity in Patients without TB, but in Whom TB was Initially Suspected?

We determined specificity estimates for a subgroup of studies involving healthy participants [40,42,44,45,57,63] compared to patients in whom TB disease was initially considered and subsequently ruled out (nontuberculous respiratory disease) [38,39,41,46,47,49–56,58,60,61,64]. In healthy volunteers, specificity values ranged from 86% to 100%, with seven (88%) studies [40,42,44,57,63] demonstrating specificity values higher than 90%. In comparison, in studies of patients with nontuberculous respiratory disease (23 studies), specificity values ranged from 47% to 100%; specificity values were higher than 90% in only seven (30%)

studies [38,41,49,51,56] (Figure 7). Compared with studies involving patients with nontuberculous respiratory disease, studies in healthy volunteers showed a higher value for the AUC (0.91 [95% CI 0.73–1.10] versus 0.87 [95% CI 0.82–0.92]), although this difference was not statistically significant, based on overlapping CIs (Figure S6A and S6B).
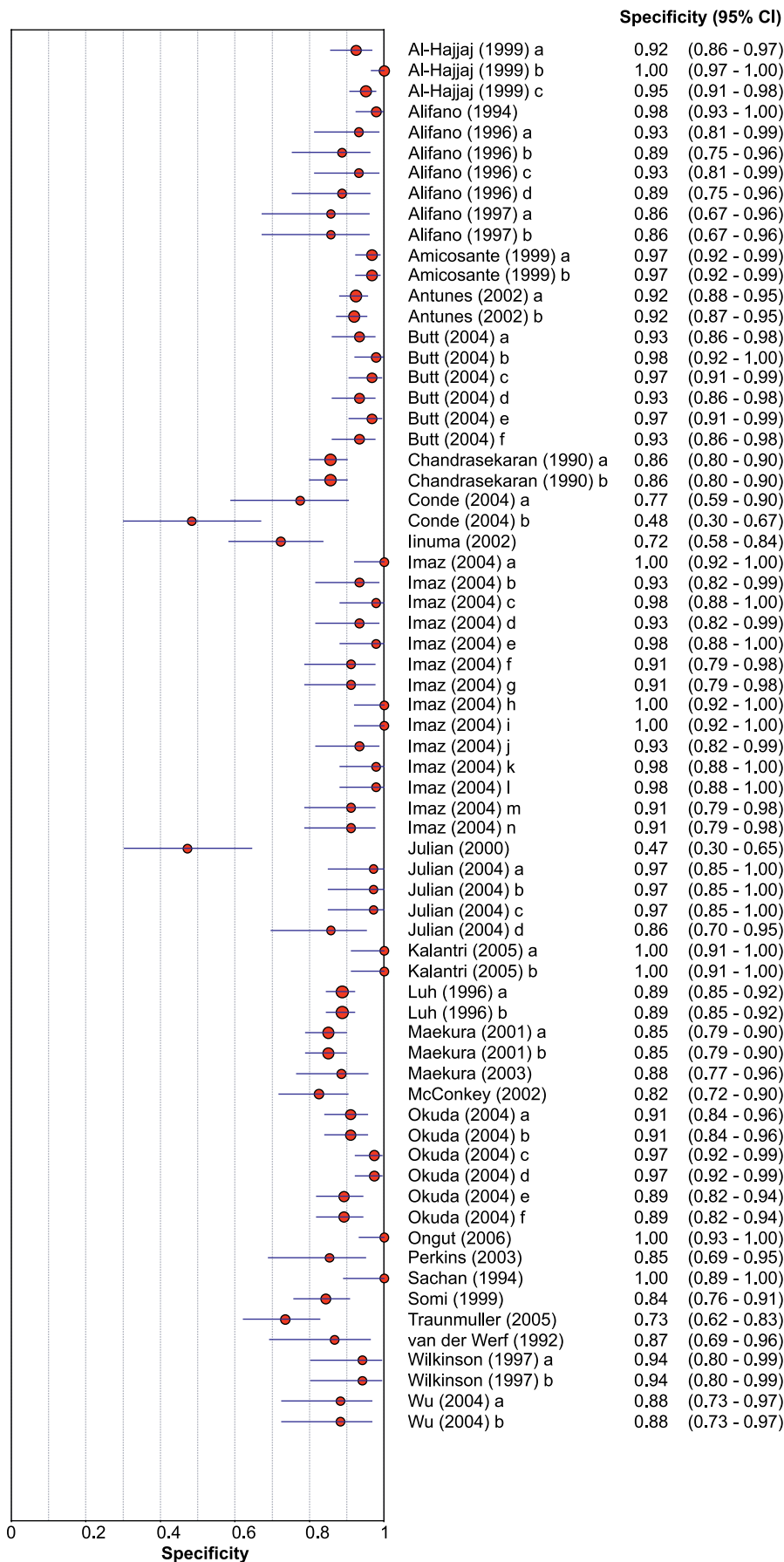
Discussion

Principal Findings

Our systematic review of 68 studies examining the performance of commercial tests for the diagnosis of pulmonary TB suggests that (1) overall, commercial tests vary widely in performance; (2) sensitivity is higher in smear-positive than smear-negative samples; (3) in studies of smear-positive patients, Anda-TB IgG by ELISA shows limited sensitivity (range 63% to 85%) and inconsistent specificity (range 73% to 100%); (4) specificity is higher in healthy volunteers than in patients in whom TB disease is initially suspected and subsequently ruled out; and (5) there are insufficient data to determine the accuracy of most commercial tests in smear-negative patients, as well as their performance in children or persons with HIV infection.

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 sensitivity, specificity and study quality in 15% of the included studies. We contacted authors for missing data. None of the studies in the review used the result from the antibody test as a reference to confirm TB (incorporation bias). When possible, we selected studies using patients with disease, in preference to healthy control participants, to evaluate how well antibody tests performed in patients with and without pulmonary TB among persons suspected of having TB. Finally, we analyzed data within specific subgroups to lessen the effect of heterogeneity.

This review also had limitations. Except for Anda-TB IgG, there were not sufficient relevant studies for specific commercial tests to provide meaningful summary measures of performance, particularly in smear-negative samples. In addition, there were no studies that met our inclusion criteria for the diagnosis of pulmonary TB in children or in patients with HIV infection. Another problem concerned limited information on clinical status and disease severity in the populations tested. Differing criteria for patient selection and greater duration and severity of illness of the study populations may have introduced significant variability in findings among studies (selection bias). Moreover, different settings (e.g., areas of high HIV prevalence, hospitals, and outpatient clinics) may have accounted for some of the observed variation in test performance. The condition of specimens may have affected the sensitivity results. Only one study [60] used fresh serum; the vast majority (87%) of studies used frozen sera; thus samples were subjected to an unknown freeze–thaw cycle history, which may decrease sensitivity [58]. A majority (75%) of studies used mycobacterial culture as the reference standard. Although in TB diagnostic trials, mycobacterial culture is considered the gold standard, culture does not detect all cases of TB; therefore, some degree of

	Specificity (95% CI)	
Al-Hajjaj (1999) a	0.92	(0.86 - 0.97)
Al-Hajjaj (1999) b	1.00	(0.97 - 1.00)
Al-Hajjaj (1999) c	0.95	(0.91 - 0.98)
Alifano (1994)	0.98	(0.93 - 1.00)
Alifano (1996) a	0.93	(0.81 - 0.99)
Alifano (1996) b	0.89	(0.75 - 0.96)
Alifano (1996) c	0.93	(0.81 - 0.99)
Alifano (1996) d	0.89	(0.75 - 0.96)
Alifano (1997) a	0.86	(0.67 - 0.96)
Alifano (1997) b	0.86	(0.67 - 0.96)
Amicosante (1999) a	0.97	(0.92 - 0.99)
Amicosante (1999) b	0.97	(0.92 - 0.99)
Antunes (2002) a	0.92	(0.88 - 0.95)
Antunes (2002) b	0.92	(0.87 - 0.95)
Butt (2004) a	0.93	(0.86 - 0.98)
Butt (2004) b	0.98	(0.92 - 1.00)
Butt (2004) c	0.97	(0.91 - 0.99)
Butt (2004) d	0.93	(0.86 - 0.98)
Butt (2004) e	0.97	(0.91 - 0.99)
Butt (2004) f	0.93	(0.86 - 0.98)
Chandrasekaran (1990) a	0.86	(0.80 - 0.90)
Chandrasekaran (1990) b	0.86	(0.80 - 0.90)
Conde (2004) a	0.77	(0.59 - 0.90)
Conde (2004) b	0.48	(0.30 - 0.67)
Iinuma (2002)	0.72	(0.58 - 0.84)
Imaz (2004) a	1.00	(0.92 - 1.00)
Imaz (2004) b	0.93	(0.82 - 0.99)
Imaz (2004) c	0.98	(0.88 - 1.00)
Imaz (2004) d	0.93	(0.82 - 0.99)
Imaz (2004) e	0.98	(0.88 - 1.00)
Imaz (2004) f	0.91	(0.79 - 0.98)
Imaz (2004) g	0.91	(0.79 - 0.98)
Imaz (2004) h	1.00	(0.92 - 1.00)
Imaz (2004) i	1.00	(0.92 - 1.00)
Imaz (2004) j	0.93	(0.82 - 0.99)
Imaz (2004) k	0.98	(0.88 - 1.00)
Imaz (2004) l	0.98	(0.88 - 1.00)
Imaz (2004) m	0.91	(0.79 - 0.98)
Imaz (2004) n	0.91	(0.79 - 0.98)
Julian (2000)	0.47	(0.30 - 0.65)
Julian (2004) a	0.97	(0.85 - 1.00)
Julian (2004) b	0.97	(0.85 - 1.00)
Julian (2004) c	0.97	(0.85 - 1.00)
Julian (2004) d	0.86	(0.70 - 0.95)
Kalantri (2005) a	1.00	(0.91 - 1.00)
Kalantri (2005) b	1.00	(0.91 - 1.00)
Luh (1996) a	0.89	(0.85 - 0.92)
Luh (1996) b	0.89	(0.85 - 0.92)
Maekura (2001) a	0.85	(0.79 - 0.90)
Maekura (2001) b	0.85	(0.79 - 0.90)
Maekura (2003)	0.88	(0.77 - 0.96)
McConkey (2002)	0.82	(0.72 - 0.90)
Okuda (2004) a	0.91	(0.84 - 0.96)
Okuda (2004) b	0.91	(0.84 - 0.96)
Okuda (2004) c	0.97	(0.92 - 0.99)
Okuda (2004) d	0.97	(0.92 - 0.99)
Okuda (2004) e	0.89	(0.82 - 0.94)
Okuda (2004) f	0.89	(0.82 - 0.94)
Ongut (2006)	1.00	(0.93 - 1.00)
Perkins (2003)	0.85	(0.69 - 0.95)
Sachan (1994)	1.00	(0.89 - 1.00)
Somi (1999)	0.84	(0.76 - 0.91)
Traunmuller (2005)	0.73	(0.62 - 0.83)
van der Werf (1992)	0.87	(0.69 - 0.96)
Wilkinson (1997) a	0.94	(0.80 - 0.99)
Wilkinson (1997) b	0.94	(0.80 - 0.99)
Wu (2004) a	0.88	(0.73 - 0.97)
Wu (2004) b	0.88	(0.73 - 0.97)

0   0.2   0.4   0.6   0.8   1
Specificity

Figure 3. Specificity Estimates of Commercial Tests for the Diagnosis of Pulmonary TB

The circles and lines represent the point estimates and 95% CIs, respectively. The size of the circle indicates the study size.
doi:10.1371/journal.pmed.0040202.g003
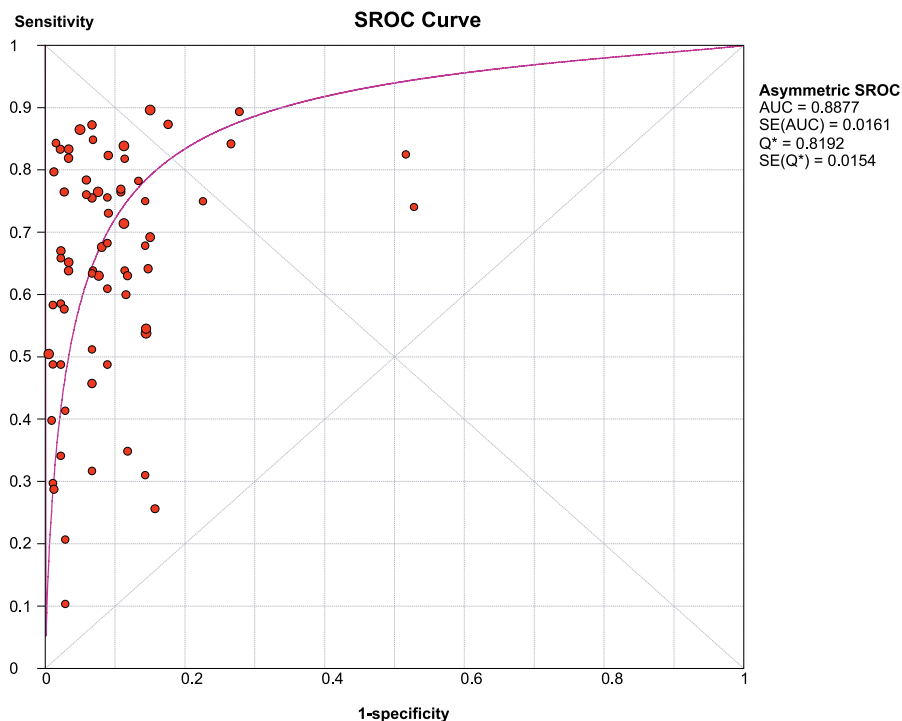
Figure 4. SROC Curve of Commercial Tests for the Diagnosis of Pulmonary TB

Each solid circle represents an individual study in the meta-analysis. The curve is the regression line that summarizes the overall diagnostic accuracy. SE (AUC), standard error of AUC; Q*, an index defined by the point on the SROC curve where the sensitivity and specificity are equal; SE (Q*), standard error of Q* index.

doi:10.1371/journal.pmed.0040202.g004

misclassification of disease for study participants was possible. We included studies from endemic countries that used sputum smear microscopy, a test with modest sensitivity (up to 70%–80%), as the reference standard; use of an insensitive reference standard may have led to biased estimates of commercial test accuracy [65]. The choice of the reference standard (culture and/or smear) for this review was a problem for pediatric TB. In this group, it is difficult to diagnose pulmonary TB on the basis of bacteriologic confirmation [9].

Another set of problems involved shortcomings in study design and quality. Only 24 (35%) studies recruited participants in a random or consecutive manner. Therefore, most studies lacked the sound probabilistic sampling framework possible in consecutive or random sampling designs. Only approximately half of the studies (31 [46%]) reported blinded interpretation of the results of the commercial test and the reference standard, a major limitation of the currently available literature. Lack of blinding may have resulted in an overestimation of the sensitivity of the commercial test result [66]. In 29 (43%) studies, different diagnostic tests were performed in TB patients and control participants (mycobacterial culture for patients and chest radiographs for control participants [verification bias]), or information about the tests to rule out TB, if performed, was not reported. Only 17 (25%) studies met all four criteria for good quality. 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 trials, such techniques have not been adequately evaluated for diagnostic data [67]. 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 publications. This represented approximately 20% of all excluded studies.

Finally, because the antigen content of commercial tests is sometimes considered proprietary information, we could not determine the antigen composition for all tests in the review. Test manufacturers may change their product without forewarning and, in some countries, such as Brazil, products are registered in the distributor's name; a change in distributor will therefore result in a change in product name, making it difficult to ascertain which tests are in current use. These points argue for independent evaluations of new immune-based diagnostic tests and, by extension, greater regulatory oversight of diagnostic tests in general, particularly for diseases that have a significant public health impact.
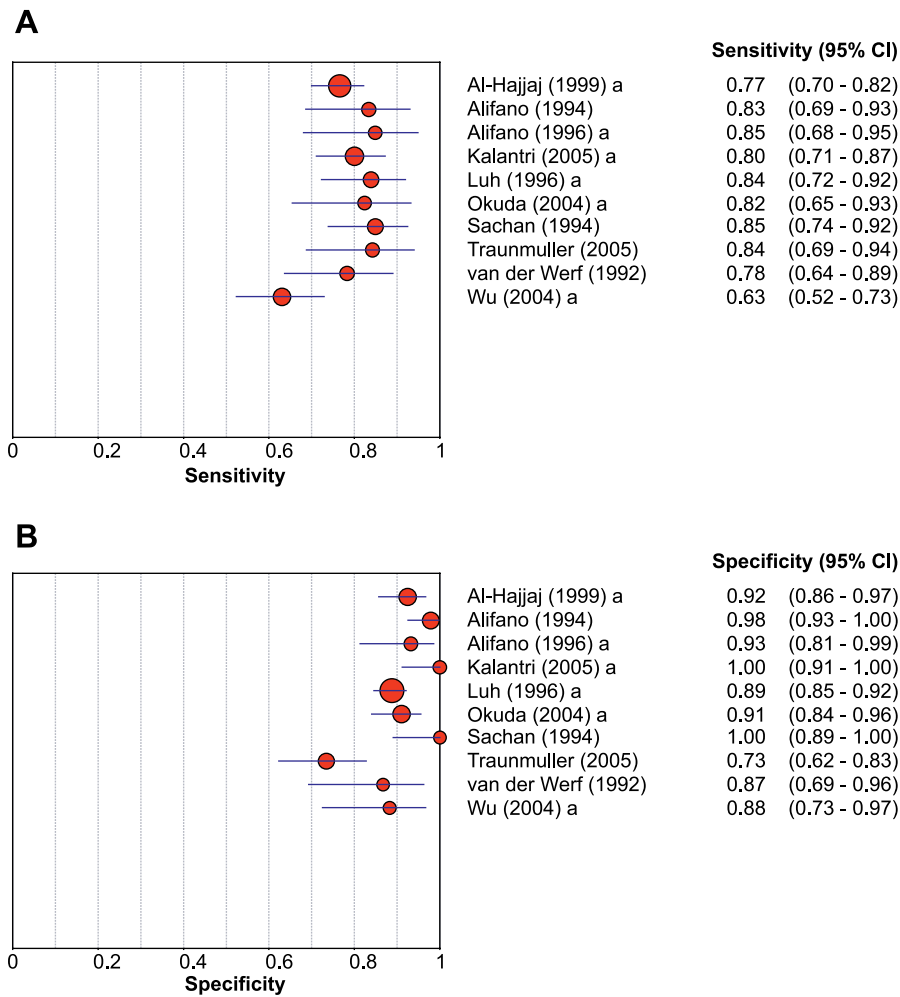
Despite the long history of failed attempts to develop an accurate serodiagnostic test for TB and the inadequate performance of the current commercial tests, it is important to emphasize that studies during the past decade have not only permitted an understanding of the lacunae in the efforts so far [19,20,68,69], but have also identified several promising candidate antigens [19,20,68–73]. A systematic analysis of the humoral immune responses of TB patients has shown that the profile of antigenic proteins of *M. tuberculosis* recognized by antibodies differs at different stages of infection and disease progression [68,69,74,75] and thus, an accurate diagnostic test for TB will need to be based on a combination of antigens [68,69,75]. In HIV-infected TB patients, several antigens that are recognized by antibodies have been delineated, however,
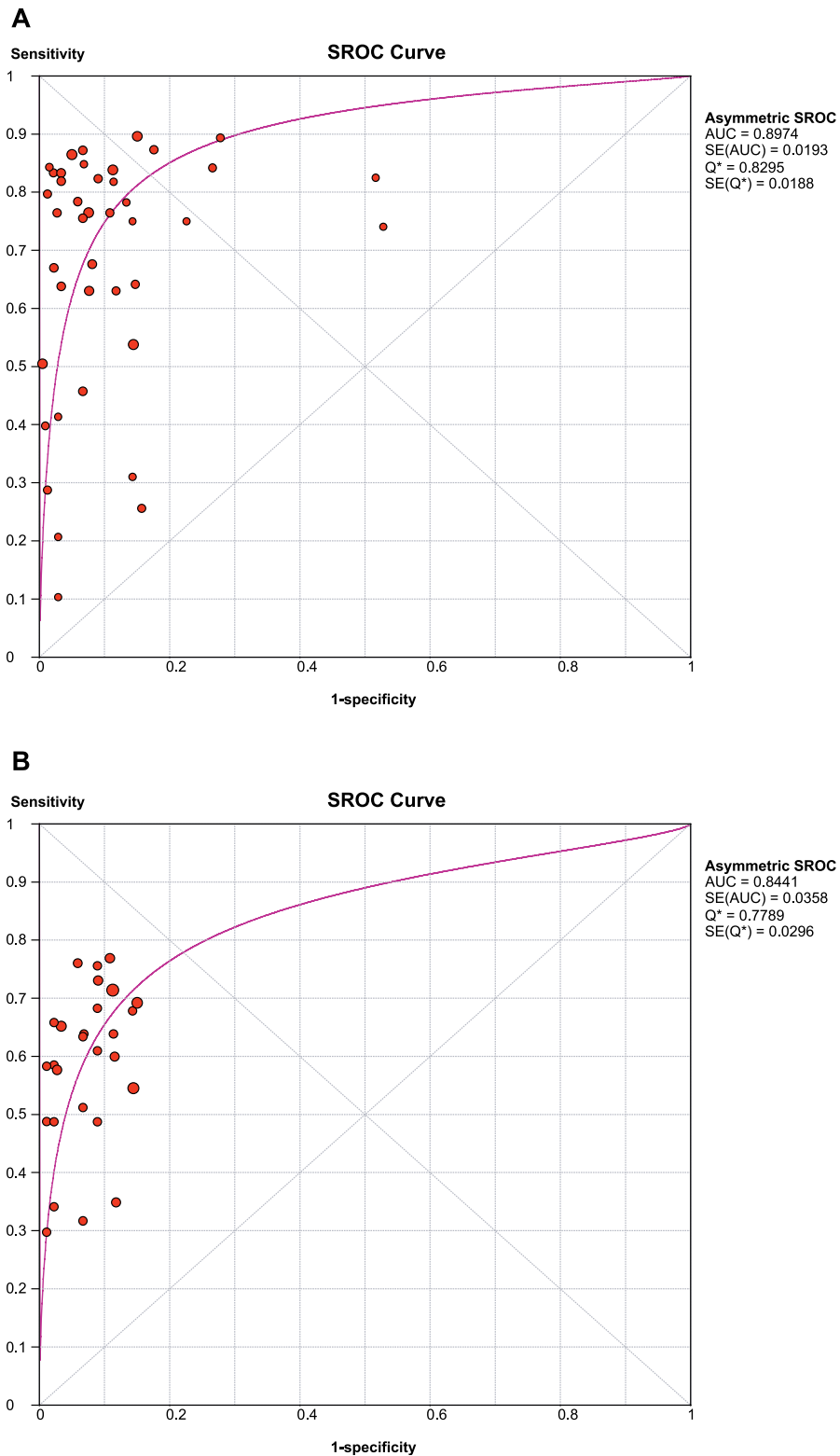
Figure 5. SROC Curve of Commercial Tests for the Diagnosis of Pulmonary TB

(A) Smear microscopy–positive patients; (B) smear microscopy–negative patients. Each solid circle represents an individual study in the meta-analysis. The curve is the regression line that summarizes the overall diagnostic accuracy. SE (AUC), standard error of AUC; Q*, an index defined by the point on the SROC curve where the sensitivity and specificity are equal; SE (Q*), standard error of Q* index.

doi:10.1371/journal.pmed.0040202.g005

A

	Sensitivity (95% CI)	
Al-Hajjaj (1999) a	0.77	(0.70 - 0.82)
Alifano (1994)	0.83	(0.69 - 0.93)
Alifano (1996) a	0.85	(0.68 - 0.95)
Kalantri (2005) a	0.80	(0.71 - 0.87)
Luh (1996) a	0.84	(0.72 - 0.92)
Okuda (2004) a	0.82	(0.65 - 0.93)
Sachan (1994)	0.85	(0.74 - 0.92)
Traunmuller (2005)	0.84	(0.69 - 0.94)
van der Werf (1992)	0.78	(0.64 - 0.89)
Wu (2004) a	0.63	(0.52 - 0.73)

B

	Specificity (95% CI)	
Al-Hajjaj (1999) a	0.92	(0.86 - 0.97)
Alifano (1994)	0.98	(0.93 - 1.00)
Alifano (1996) a	0.93	(0.81 - 0.99)
Kalantri (2005) a	1.00	(0.91 - 1.00)
Luh (1996) a	0.89	(0.85 - 0.92)
Okuda (2004) a	0.91	(0.84 - 0.96)
Sachan (1994)	1.00	(0.89 - 1.00)
Traunmuller (2005)	0.73	(0.62 - 0.83)
van der Werf (1992)	0.87	(0.69 - 0.96)
Wu (2004) a	0.88	(0.73 - 0.97)

Figure 6. Sensitivity and Specificity Estimates of Anda-TB IgG for the Diagnosis of Pulmonary TB, Smear Microscopy–Positive Patients
(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.
doi:10.1371/journal.pmed.0040202.g006

only the 81-kDa malate synthase protein has been evaluated in patients from different countries [19,20,70–73]. It is encouraging that this antigen provided equivalent sensitivity in patients from India, the United States, and Uganda [19,20,70] since it has been suggested that differences in genetic make-up might account for the observed variations in antibody responses to specific *M. tuberculosis* antigens [13,76].

Antigens that are immunodominant during extrapulmonary and pediatric TB also need to be identified. A systematic review of 21 studies evaluating the performance of commercial tests for the diagnosis of extrapulmonary TB showed highly variable estimates of sensitivity and specificity, and no studies were of sufficient quality to enable their evaluation in patients with HIV infection or in children—the two groups for which the test would be most useful [77].
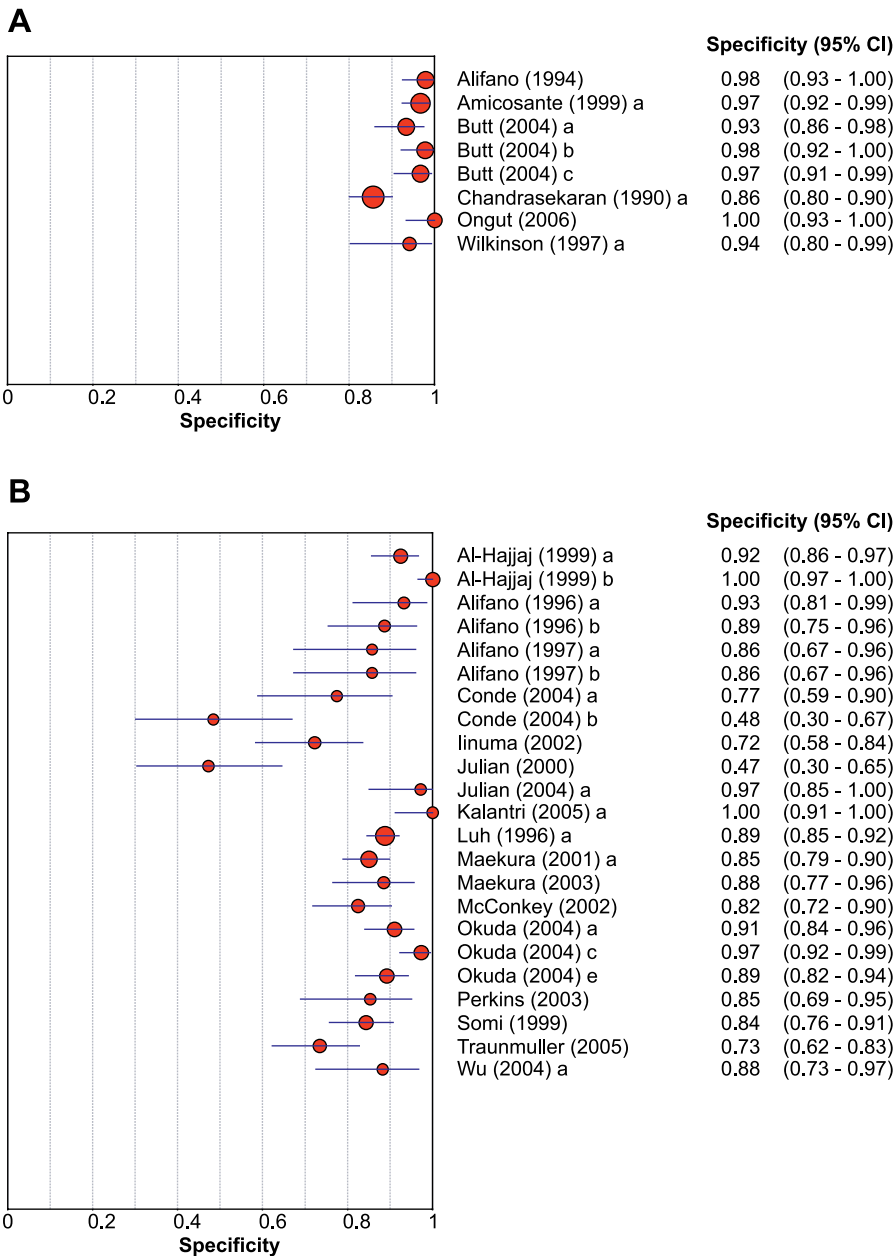
Conclusions and Policy Implications

The evidence provided in this systematic review suggests that, at this point in time, published data on commercial antibody detection tests produce inconsistent estimates of accuracy, and none of the assays perform well enough to replace sputum smear microscopy. These tests thus have little or no role to play in the diagnosis of pulmonary TB at the present time. Given these findings, we express concern that commercial tests in present use may divert resources in developing countries away from smear microscopy. Our findings underscore the need for greater regulatory oversight of in vitro diagnostics and improved capacity in countries to design, conduct, and report diagnostic test evaluations which, in turn, can guide procurement and clinical practice. Lack of methodological rigor in these studies was a cause for concern. Recent articles have attested to the mediocre quality of diagnostic studies for TB [26,78]. Use of guidelines such as the Standards for Reporting of Diagnostic Accuracy (STARD) [79] and the tool for quality assessment of diagnostic accuracy studies (the QUADAS tool) [80] 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 [81].

It is important that the literature from research laboratories that have evaluated the immunodiagnostic potential of different antigens be reviewed to determine whether there are useful antigens which have been described but whose potential has not been fully exploited. The increased understanding of the humoral immune responses in TB patients and the new tools of genomics and proteomics could well lead

A

Study	Specificity (95% CI)
Alifano (1994)	0.98 (0.93 - 1.00)
Amicosante (1999) a	0.97 (0.92 - 0.99)
Butt (2004) a	0.93 (0.86 - 0.98)
Butt (2004) b	0.98 (0.92 - 1.00)
Butt (2004) c	0.97 (0.91 - 0.99)
Chandrasekaran (1990) a	0.86 (0.80 - 0.90)
Ongut (2006)	1.00 (0.93 - 1.00)
Wilkinson (1997) a	0.94 (0.80 - 0.99)

B

Study	Specificity (95% CI)
Al-Hajjaj (1999) a	0.92 (0.86 - 0.97)
Al-Hajjaj (1999) b	1.00 (0.97 - 1.00)
Alifano (1996) a	0.93 (0.81 - 0.99)
Alifano (1996) b	0.89 (0.75 - 0.96)
Alifano (1997) a	0.86 (0.67 - 0.96)
Alifano (1997) b	0.86 (0.67 - 0.96)
Conde (2004) a	0.77 (0.59 - 0.90)
Conde (2004) b	0.48 (0.30 - 0.67)
Iinuma (2002)	0.72 (0.58 - 0.84)
Julian (2000)	0.47 (0.30 - 0.65)
Julian (2004) a	0.97 (0.85 - 1.00)
Kalantri (2005) a	1.00 (0.91 - 1.00)
Luh (1996) a	0.89 (0.85 - 0.92)
Maekura (2001) a	0.85 (0.79 - 0.90)
Maekura (2003)	0.88 (0.77 - 0.96)
McConkey (2002)	0.82 (0.72 - 0.90)
Okuda (2004) a	0.91 (0.84 - 0.96)
Okuda (2004) c	0.97 (0.92 - 0.99)
Okuda (2004) e	0.89 (0.82 - 0.94)
Perkins (2003)	0.85 (0.69 - 0.95)
Somi (1999)	0.84 (0.76 - 0.91)
Traunmuller (2005)	0.73 (0.62 - 0.83)
Wu (2004) a	0.88 (0.73 - 0.97)

Figure 7. Specificity Estimates of Commercial Tests
(A) Healthy control participants; (B) patients with nontuberculous respiratory disease. Studies using identical comparison groups appear only once. The circles and lines represent the point estimates and 95% CIs, respectively. The size of the circle indicates the study size.
doi:10.1371/journal.pmed.0040202.g007

to devising the simple, rapid, and accurate serodiagnostic test that that has eluded us so far. Activities aimed at discovering new antigens with immunodiagnostic potential need to be intensified. Trials of new serodiagnostic tests for pulmonary TB must adequately address the particular challenges presented by smear microscopy–negative patients, children, and persons with HIV infection.

Supporting Information

Figure S1. Sensitivity Estimates of Commercial Tests for the Diagnosis of Pulmonary TB, Smear Microscopy–Positive Patients

The circles and lines represent the point estimates and 95% CIs, respectively. The size of the circle indicates the study size. Numbers in parentheses indicate references. EIA, enzyme immunoassay; IgG, IgM, IgA (G, M, A), immunoglobulin G, M, A, respectively; KAT, kaolin agglutination test; P Plus, Pathozyme TB Complex Plus; Path, Pathozyme; TBGL, tuberculosis glycolipid assay.

Found at doi:10.1371/journal.pmed.0040202.sg001 (262 KB PDF).

Figure S2. Specificity Estimates of Commercial Tests for the Diagnosis of Pulmonary TB, Smear Microscopy–Positive Patients

The circles and lines represent the point estimates and 95% CIs, respectively. The size of the circle indicates the study size. Numbers in parentheses indicate references. EIA, enzyme immunoassay; IgG, IgM, IgA (G, M, A), immunoglobulin G, M, A, respectively; KAT, kaolin agglutination test; P Plus, Pathozyme TB Complex Plus; Path, Pathozyme; TBGL, tuberculosis glycolipid assay.

Found at doi:10.1371/journal.pmed.0040202.sg002 (257 KB PDF).

Figure S3. Sensitivity Estimates of Commercial Tests for the Diagnosis of Pulmonary TB, Smear Microscopy–Negative Patients

The circles and lines represent the point estimates and 95% CIs, respectively. The size of the circle indicates the study size. Numbers in parentheses indicate references. EIA, enzyme immunoassay; IgG, IgM, IgA (G, M, A), immunoglobulin G, M, A, respectively; KAT, kaolin agglutination test; P Plus, Pathozyme TB Complex Plus; Path, Pathozyme; TBGL, tuberculosis glycolipid assay.

Found at doi:10.1371/journal.pmed.0040202.sg003 (259 KB PDF).

Figure S4. Specificity Estimates of Commercial Tests for the Diagnosis of Pulmonary TB, Smear Microscopy–Negative Patients

The circles and lines represent the point estimates and 95% CIs, respectively. The size of the circle indicates the study size. Numbers in parentheses indicate references. EIA, enzyme immunoassay; IgG, IgM, IgA (G, M, A), immunoglobulin G, M, A, respectively; KAT, kaolin agglutination test; P Plus, Pathozyme TB Complex Plus; Path, Pathozyme; TBGL, tuberculosis glycolipid assay.

Found at doi:10.1371/journal.pmed.0040202.sg004 (256 KB PDF).

Figure S5. SROC Curve of Anda-TB IgG for the Diagnosis of Pulmonary TB, Smear Microscopy–Positive Patients

Each solid circle represents an individual study in the meta-analysis. The curve is the regression line that summarizes the overall diagnostic accuracy. SE (AUC), standard error of AUC; Q*, an index defined by the point on the SROC curve where the sensitivity and specificity are equal; SE (Q*), standard error of Q* index.

Found at doi:10.1371/journal.pmed.0040202.sg005 (234 KB PDF).

Figure S6. SROC Curve of Commercial Tests for the Diagnosis of Pulmonary TB

(A) Healthy control participants; (B) patients with nontuberculous respiratory disease. Each solid circle represents an individual study in the meta-analysis. The curve is the regression line that summarizes the overall diagnostic accuracy. SE (AUC), standard error of AUC; Q*, an index defined by the point on the SROC curve where the sensitivity and specificity are equal; SE (Q*), standard error of Q* index.

Found at doi:10.1371/journal.pmed.0040202.sg006 (266 KB PDF).

Acknowledgments

References

1. World Health Organization (2007) Global tuberculosis control: Surveillance, planning, financing. WHO report. Geneva: World Health Organization. WHO/HTM/TB/2007.376.
2. Dye C, Watt CJ, Bleed DM, Hosseini SM, Raviglione MC (2005) Evolution of tuberculosis control and prospects for reducing tuberculosis incidence, prevalence, and deaths globally. JAMA 293: 2767–2775.
3. Toman K (2004) What are the advantages and disadvantages of fluorescence microscopy? In: Frieden T, editor. Toman's tuberculosis: Case detection, treatment, and monitoring—questions and answers. 2nd edition. Geneva: World Health Organization. pp. 31–34.
4. Steingart KR, Ng V, Henry M, Hopewell PC, Ramsay A, et al. (2006) Sputum processing methods to improve the sensitivity of smear microscopy for tuberculosis: A systematic review. Lancet Infect Dis 6: 664–674.
5. Behr MA, Warren SA, Salamon H, Hopewell PC, Ponce de Leon A, et al. (1999) Transmission of *Mycobacterium tuberculosis* from patients smear-negative for acid-fast bacilli. Lancet 353: 444–449.
6. Grzybowski S, Barnett GD, Styblo K (1975) Contacts of cases of active pulmonary tuberculosis. Bull Int Union Tuberc 50: 90–106.
7. Urbanczik R (1985) Present position of microscopy and of culture in diagnostic mycobacteriology. Zentralbl Bakteriol Mikrobiol Hyg [A] 260: 81–87.

8. Perkins MD, Roscigno G, Zumla A (2006) Progress towards improved tuberculosis diagnostics for developing countries. Lancet 367: 942–943.
9. Shingadia D, Novelli V (2003) Diagnosis and treatment of tuberculosis in children. Lancet Infect Dis 3: 624–632.
10. Elliott AM, Namaambo K, Allen BW, Luo N, Hayes RJ, et al. (1993) Negative sputum smear results in HIV-positive patients with pulmonary tuberculosis in Lusaka, Zambia. Tuberc Lung Dis 74: 191–194.
11. Johnson JL, Vjecha MJ, Okwera A, Hatanga E, Byekwaso F, et al. (1998) Impact of human immunodeficiency virus type-1 infection on the initial bacteriologic and radiographic manifestations of pulmonary tuberculosis in Uganda. Makerere University-Case Western Reserve University Research Collaboration. Int J Tuberc Lung Dis 2: 397–404.
12. Menzies D (2004) What is the current and potential role of diagnostic tests other than sputum microscopy and culture? In: Frieden T, editor. Toman's tuberculosis: Case detection, treatment, and monitoring—Questions and answers. 2nd edition. Geneva: World Health Organization. pp. 87–91.
13. Bothamley GH (1995) Serological diagnosis of tuberculosis. Eur Respir J Suppl 20: 676S–688S.
14. Chan ED, Heifets L, Iseman MD (2000) Immunologic diagnosis of tuberculosis: A review. Tuberc Lung Dis 80: 131–140.
15. Daniel TM (1989) Rapid diagnosis of tuberculosis: Laboratory techniques applicable in developing countries. Rev Infect Dis 11: S471–S478.
16. Daniel TM (1990) The rapid diagnosis of tuberculosis: A selective review. J Lab Clin Med 116: 277–282.
17. Gennaro ML (2000) Immunologic diagnosis of tuberculosis. Clin Infect Dis 30: S243–S246.
18. Iseman MD (2000) Immunity and pathogenesis. In: Iseman MD, editor. A clinician's guide to tuberculosis. Philadelphia: Lippincott Williams and Wilkins. pp. 63–96.
19. Laal S (2004) Immunodiagnosis. In: Rom WN, Garay SM, editors. Tuberculosis. Philadelphia: Lippincott Williams and Wilkins. pp. 185–191.
20. Laal S, Skeiky YA (2005) Immune-based methods. In: Cole ST, editor. Tuberculosis and the tubercle bacillus. Washington (D. C.): ASM Press. pp. 71–83.
21. Palomino JC (2005) Nonconventional and new methods in the diagnosis of tuberculosis: Feasibility and applicability in the field. Eur Respir J 26: 339–350.
22. Pai M, Kalantri S, Dheda K (2006) New tools and emerging technologies for the diagnosis of tuberculosis: Part II. Active tuberculosis and drug resistance. Expert Rev Mol Diagn 6: 423–432.
23. Boehme C, Molokova E, Minja F, Geis S, Loscher T, et al. (2005) Detection of mycobacterial lipoarabinomannan with an antigen-capture ELISA in unprocessed urine of Tanzanian patients with suspected tuberculosis. Trans R Soc Trop Med Hyg 99: 893–900.
24. Tessema TA, Hamasur B, Bjun G, Svenson S, Bjorvatn B (2001) Diagnostic evaluation of urinary lipoarabinomannan at an Ethiopian tuberculosis centre. Scand J Infect Dis 33: 279–284.
25. Kunter E, Cerrahoglu K, Ilvan A, Isitmangil T, Turken O, et al. (2003) The value of pleural fluid anti-A60 IgM in BCG-vaccinated tuberculous pleurisy patients. Clin Microbiol Infect 9: 212–220.
26. Small PM, Perkins MD (2000) More rigour needed in trials of new diagnostic agents for tuberculosis. Lancet 356: 1048–1049.
27. Walsh A, McNerney R (2004) Guidelines for establishing trials of new tests to diagnose tuberculosis in endemic countries. Int J Tuberc Lung Dis 8: 609–613.
28. Special Programme for Research and Training in Tropical Diseases, Foundation for Innovative New Diagnostics (2006) Current TB diagnostics market. Diagnostics for tuberculosis: Global demand and market potential. Geneva: World Health Organization. pp. 49–72.
29. Janeway CA, Travers P, Walport M, Shlomchik M (2005) Appendix I, Immunologists' toolbox. Immunobiology, the immune system in health and disease. 6th edition. New York: Garland Science Publishing. pp. 683–729.
30. Bangs Laboratories Immunochromatographic, lateral flow or strip tests development ideas. Available: http://www.pall.com/34445__4154.asp#4166. Accessed 10 May 2007.
31. Pai M, McCulloch M, Enanoria W, Colford JM Jr (2004) Systematic reviews of diagnostic test evaluations: What's behind the scenes? ACP J Club 141: A11–A13.
32. Littenberg B, Moses LE (1993) Estimating diagnostic accuracy from multiple conflicting reports: A new meta-analytic method. Med Decis Making 13: 313–321.
33. SPSS (2006) SPSS for Windows. 14.0.1.366 edition. Chicago: SPSS.
34. Zamora J, Abraira V, Muriel A, Khan KS, Coomarasamy A (2006) Meta-DiSc: A software for meta-analysis of test accuracy data. BMC Med Res Methodol 6: 31.
35. Walter SD (2002) Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. Stat Med 21: 1237–1256.
36. Jones CM, Athanasiou T (2005) Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. Ann Thorac Surg 79: 16–20.
37. Lijmer JG, Bossuyt PM, Heisterkamp SH (2002) Exploring sources of heterogeneity in systematic reviews of diagnostic tests. Stat Med 21: 1525–1537.
38. Al-Hajjaj MS, Gad-el-Rab MO, Al-Orainey IO, Al-Kassimi FA (1999)

Improved sensitivity for detection of tuberculosis cases by a modified Anda-TB ELISA test. Tuberc Lung Dis 79: 181–185.

39. Alifano M, De Pascalis R, Sofia M, Faraone S, Del Pezzo M, et al. (1997) Evaluation of IgA-mediated humoral immune response against the mycobacterial antigen P-90 in diagnosis of pulmonary tuberculosis. Chest 111: 601–605.

40. Alifano M, Del Pezzo M, Lamberti C, Faraone S, Covelli I (1994) ELISA method for evaluation of anti-A60 IgG in patients with pulmonary and extrapulmonary tuberculosis. New Microbiol 17: 37–43.

41. Alifano M, Sofia M, Mormile M, Micco A, Mormile AF, et al. (1996) IgA immune response against the mycobacterial antigen A60 in patients with active pulmonary tuberculosis. Respiration 63: 292–297.

42. Amicosante M, Houde M, Guaraldi G, Saltini C (1999) Sensitivity and specificity of a multi-antigen ELISA test for the serological diagnosis of tuberculosis. Int J Tuberc Lung Dis 3: 736–740.

43. Antunes A, Nina J, David S (2002) Serological screening for tuberculosis in the community: An evaluation of the Mycodot procedure in an African population with high HIV-2 prevalence (Republic of Guinea-Bissau). Res Microbiol 153: 301–305.

44. Butt T, Malik HS, Abbassi SA, Ahmad RN, Mahmood A, et al. (2004) Genus and species-specific IgG and IgM antibodies for pulmonary tuberculosis. J Coll Physicians Surg Pak 14: 105–107.

45. Chandrasekaran S, Gupta EVV, Chauhan MM, Baily GVJ, Chaudhuri K (1990) Serodiagnosis of pulmonary tuberculosis by kaolin agglutination test. Indian J Tuberc 37: 11–15.

46. Conde MB, Suffys P, Silva J, Kritski AL, Dorman SE (2004) Immunoglobulin A (IgA) and IgG immune responses against P-90 antigen for diagnosis of pulmonary tuberculosis and screening for Mycobacterium tuberculosis infection. Clin Diagn Lab Immunol 11: 94–97.

47. Iinuma Y, Senda K, Takakura S, Ichiyama S, Tano M, et al. (2002) Evaluation of a commercially available serologic assay for antibodies against tuberculosis-associated glycolipid antigen. Clin Chem Lab Med 40: 832–836.

48. Imaz MS, Comini MA, Zerbini E, Sequeira MD, Latini O, et al. (2004) Evaluation of commercial enzyme-linked immunosorbent assay kits for detection of tuberculosis in Argentinean population. J Clin Microbiol 42: 884–887.

49. Julian E, Matas L, Alcaide J, Luquin M (2004) Comparison of antibody responses to a potential combination of specific glycolipids and proteins for test sensitivity improvement in tuberculosis serodiagnosis. Clin Diagn Lab Immunol 11: 70–76.

50. Julian E, Matas L, Hernandez A, Alcaide J, Luquin M (2000) Evaluation of a new serodiagnostic tuberculosis test based on immunoglobulin A detection against Kp-90 antigen. Int J Tuberc Lung Dis 4: 1082–1085.

51. Kalantri Y, Hemvani N, Bhatia GC, Chitnis DS (2005) ELISA kit evaluation for IgG and IgM antibodies to A-60 tubercular protein antigen. Indian J Med Sci 59: 337–346.

52. Luh KT, Yu CJ, Yang PC, Lee LN (1996) Tuberculosis antigen A60 serodiagnosis in tuberculous infection: Application in extrapulmonary and smear-negative pulmonary tuberculosis. Respirology 1: 145–151.

53. Maekura R, Kohno H, Hirotani A, Okuda Y, Ito M, et al. (2003) Prospective clinical evaluation of the serologic tuberculous glycolipid test in combination with the nucleic acid amplification test. J Clin Microbiol 41: 1322–1325.

54. Maekura R, Okuda Y, Nakagawa M, Hiraga T, Yokota S, et al. (2001) Clinical evaluation of anti-tuberculous glycolipid immunoglobulin G antibody assay for rapid serodiagnosis of pulmonary tuberculosis. J Clin Microbiol 39: 3603–3608.

55. McConkey SJ, Youssef FG, Azem E, Frenck RW, Weil GJ (2002) 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 6: 246–252.

56. Okuda Y, Maekura R, Hirotani A, Kitada S, Yoshimura K, et al. (2004) Rapid serodiagnosis of active pulmonary Mycobacterium tuberculosis by analysis of results from multiple antigen-specific tests. J Clin Microbiol 42: 1136–1141.

57. Ongut G, Ogunc D, Gunseren F, Ogus C, Donmez L, et al. (2006) Evaluation of the ICT tuberculosis test for the routine diagnosis of tuberculosis. BMC Infect Dis 6: 37.

58. Perkins MD, Conde MB, Martins A, Kritski AL (2003) Serologic diagnosis of tuberculosis using a simple commercial multiantigen assay. Chest 123: 107–112.

59. Sachan A, Gupta R, Gupta D, Gautam K, Gupta D (1994) Diagnostic significance of antigen A-60 by ELISA test in pulmonary tuberculosis. Indian J Tuberc 41: 239–243.

60. Somi GR, O'Brien RJ, Mfinanga GS, Ipuge YA (1999) Evaluation of the MycoDot (TM) test in patients with suspected tuberculosis in a field setting in Tanzania. Int J Tuberc Lung Dis 3: 231–238.

61. Traunmuller F, Haslinger I, Lagler H, Wolfgang G, Zeitlinger MA, et al. (2005) Influence of the washing buffer composition on the sensitivity of an enzyme-linked immunosorbent assay using mycobacterial glycolipids as capture antigens. J Immunoassay Immunochem 26: 179–188.

62. van der Werf TS, Das PK, van Soolingen D, Yong S, van der Mark TW, et al. (1992) Sero-diagnosis of tuberculosis with A60 antigen enzyme-linked immunosorbent assay: Failure in HIV-infected individuals in Ghana. Med Microbiol Immunol (Berl) 181: 71–76.

63. Wilkinson RJ, Haslov K, Rappuoli R, Giovannoni F, Narayanan PR, et al. (1997) Evaluation of the recombinant 38-kilodalton antigen of Mycobacterium tuberculosis as a potential immunodiagnostic reagent. J Clin Microbiol 35: 553–557.

64. Wu HP, Shieh WB, Hsien FK, Hua CC (2004) The significance of Mycobacterium tuberculosis antibody, antigen 60 IgG in patients with abnormal chest radiography. Chang Gung Med J 27: 869–876.

65. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, et al. (2004) Sources of variation and bias in studies of diagnostic accuracy: A systematic review. Ann Intern Med 140: 189–202.

66. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, et al. (1999) Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 282: 1061–1066.

67. Tatsioni A, Zarin DA, Aronson N, Samson DJ, Flamm CR, et al. (2005) Challenges in systematic reviews of diagnostic technologies. Ann Intern Med 142: 1048–1055.

68. Samanich K, Belisle JT, Laal S (2001) Homogeneity of antibody responses in tuberculosis patients. Infect Immun 69: 4600–4609.

69. Sartain MJ, Slayden RA, Singh KK, Laal S, Belisle JT (2006) Disease state differentiation and identification of tuberculosis biomarkers via native antigen array profiling. Mol Cell Proteomics 5: 2102–2113.

70. Hendrickson RC, Douglass JF, Reynolds LD, McNeill PD, Carter D, et al. (2000) Mass spectrometric identification of Mtb81, a novel serological marker for tuberculosis. J Clin Microbiol 38: 2354–2361.

71. Houghton RL, Lodes MJ, Dillon DC, Reynolds LD, Day CH, et al. (2002) Use of multiepitope polyproteins in serodiagnosis of active tuberculosis. Clin Diagn Lab Immunol 9: 883–891.

72. Weldingh K, Rosenkrands I, Okkels LM, Doherty TM, Andersen P (2005) Assessing the serodiagnostic potential of 35 Mycobacterium tuberculosis proteins and identification of four novel serological antigens. J Clin Microbiol 43: 57–65.

73. Mukherjee S, Daifalla N, Zhang Y, Douglass J, Brooks L, et al. (2004) Potential serological use of a recombinant protein that is a replica of a Mycobacterium tuberculosis protein found in the urine of infected mice. Clin Diagn Lab Immunol 11: 280–286.

74. Samanich KM, Belisle JT, Sonnenberg MG, Keen MA, Zolla-Pazner S, et al. (1998) Delineation of human antibody responses to culture filtrate antigens of Mycobacterium tuberculosis. J Infect Dis 178: 1534–1538.

75. Singh KK, Zhang X, Patibandla AS, Chien P Jr, Laal S (2001) Antigens of Mycobacterium tuberculosis expressed during preclinical tuberculosis: Serological immunodominance of proteins with repetitive amino acid sequences. Infect Immun 69: 4185–4191.

76. Lyashchenko K, Colangeli R, Houde M, Al Jahdali H, Menzies D, et al. (1998) Heterogeneous antibody responses in tuberculosis. Infect Immun 66: 3936–3940.

77. Steingart KR, Henry M, Laal S, Hopewell PC, Ramsay A, et al. (2007) A systematic review of commercial serological antibody detection tests for the diagnosis of extrapulmonary tuberculosis. Thorax. In press.

78. Pai M, O'Brien R (2006) Tuberculosis diagnostics trials: Do they lack methodological rigor? Expert Rev Mol Diagn 6: 509–514.

79. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, et al. (2003) Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Ann Intern Med 138: 40–44.

80. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J (2003) The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 3: 25.

81. Banoo S, Bell D, Bossuyt P, Herring A, Mabey D, et al. (2006) Evaluation of diagnostic tests for infectious diseases: General principles. Nat Rev Microbiol 4: S21–S31.

Editors' Summary

Background. Tuberculosis (TB) is, globally, one of the most important infectious diseases. It is thought that in 2005 around 1.6 million people died as a result of TB. Controlling TB requires that the disease is correctly diagnosed so that it can then be promptly treated, which will reduce the risk of infection being passed on to other individuals. The method normally used for diagnosing TB disease in poor countries (where most people with TB disease live) involves taking a sample of mucus coughed up from the lungs; this mucus is then spread thinly onto a glass slide, dyed, and viewed under the microscope. The bacteria responsible for TB take up the dye in a particular pattern and can be clearly seen under the microscope. Although this test (sputum smear) is relatively straightforward to carry out even where facilities are basic, it is not particularly good at identifying TB disease in children or amongst individuals who are HIV-positive. Finally, the sputum smear test is also not very sensitive; that is, many people who have TB disease may not give a positive reading. Therefore, there is an urgent need to develop and evaluate new tests that are suitable for use in poor countries, which will accurately diagnose TB disease, especially amongst children and people who are HIV-positive.

Why Was This Study Done? New tests for TB have become available which detect whether an individual has raised antibodies against particular proteins and other substances present on the surface of the TB bacterium. These tests are carried out on blood samples, once blood cells and other factors have been taken out. These antibody tests are often quite simple to carry out, so in principle they could be suitable for use in developing countries. Since the tests are available on the market and can be freely used in some developing countries without any need for government regulatory bodies to approve them, it is important to know how good these tests are at diagnosing TB disease. The researchers here wanted, therefore, to evaluate all of the available data relating to the accuracy of antibody detection tests for diagnosis of TB disease.

What Did the Researchers Do and Find? In order to evaluate all of the information available on commercial antibody detection tests for diagnosis of TB disease of the lungs, the researchers carried out a systematic review. First, they searched biomedical literature databases using specific terms to identify studies for inclusion. A study was included in their analysis if the commercial test was compared against one of two other standard tests (sputum smear microscopy, or growth of TB bacteria in culture). One researcher from the team then pulled out specific pieces of information from each published study: these included the type of study design; information on study participants; the type of test; what the test was compared against; and finally the results of evaluation of the test. A second researcher pulled out pieces of

information from several of the same studies. The researchers then compared the information to ensure that it was recorded correctly. Each study was also assigned a quality rating, based on four distinct criteria. For each type of test, the researchers used the data in the published studies to work out the test's accuracy, both in terms of its ability to give a positive reading for people who have TB disease as well as its ability to give a negative reading for people who do not have TB disease.

The researchers found 27 papers meeting their criteria. These papers reported the results of 68 original studies. Nine different commercial tests were examined in the studies. Overall, the studies seemed to be of relatively poor quality, with only 25% of them meeting all four of the researchers' criteria for a good-quality study. The different studies appeared to produce varying results for the accuracy of these commercial tests. In particular, the tests seemed to be less accurate at detecting TB disease amongst people who had a negative sputum smear than amongst people with a positive sputum smear. When all the data for these different studies were combined, the statistics indicated that the commercial tests, overall, were only modestly accurate for diagnosis of TB disease. None of the studies had been carried out in children or in HIV-positive people.

What Do These Findings Mean? The results of this systematic review suggest that the commercial antibody detection tests considered here are not particularly useful in diagnosis of TB disease as compared to other tests, such as sputum smear and bacterial culture. Some people are concerned that there is pressure in certain developing countries to start using these tests, but the current data do not support greater use. This systematic review also highlights the fact that many studies evaluating commercial TB tests are of poor quality, and that further research needs to be done to evaluate the accuracy of different TB tests amongst children and HIV-positive patients.

Additional Information. Please access these Web sites via the online version of this summary at http://dx.doi.org/10.1371/journal.pmed.0040202.

- World Health Organization Stop TB Department website. Information about the current Stop TB strategy, data and factsheets about TB, and other resources are available
- Questions and Answers about Tuberculosis provided by the US Centers for Disease Control and Prevention
- Information about TB tests from Médicins sans Frontières (MSF). Links to MSF reports on new diagnostic tests are also available
- Wikipedia entry on Systematic Reviews (Note: Wikipedia is an internet encyclopedia anyone can edit)