

Commercial Serological Tests for the Diagnosis of Active Pulmonary and Extrapulmonary Tuberculosis: An Updated Systematic Review and Meta-Analysis

Karen R. Steingart¹, Laura L. Flores^{2,3}, Nandini Dendukuri⁴, Ian Schiller⁴, Suman Laal^{5,6,7}, Andrew Ramsay⁸, Philip C. Hopewell^{2,3}, Madhukar Pai^{4*}

1 Department of Health Services, University of Washington School of Public Health, Seattle, Washington, United States of America, **2** Division of Pulmonary and Critical Care Medicine, San Francisco General Hospital, University of California, San Francisco, California, United States of America, **3** Curry International Tuberculosis Center, University of California, San Francisco, California, United States of America, **4** Department of Epidemiology, Biostatistics, and Occupational Health, McGill University & Montreal, Chest Institute, Montreal, Quebec, Canada, **5** Department of Pathology, New York University Langone Medical Center, New York, New York, United States of America, **6** Department of Microbiology, New York University Langone Medical Center, New York, New York, United States of America, **7** Veterans Affairs Medical Center, New York, New York, United States of America, **8** UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases, World Health Organization, Geneva, Switzerland

Abstract

Background: Serological (antibody detection) tests for tuberculosis (TB) are widely used in developing countries. As part of a World Health Organization policy process, we performed an updated systematic review to assess the diagnostic accuracy of commercial serological tests for pulmonary and extrapulmonary TB with a focus on the relevance of these tests in low- and middle-income countries.

Methods and Findings: We used methods recommended by the Cochrane Collaboration and GRADE approach for rating quality of evidence. In a previous review, we searched multiple databases for papers published from 1 January 1990 to 30 May 2006, and in this update, we add additional papers published from that period until 29 June 2010. We prespecified subgroups to address heterogeneity and summarized test performance using bivariate random effects meta-analysis. For pulmonary TB, we included 67 studies (48% from low- and middle-income countries) with 5,147 participants. For all tests, estimates were variable for sensitivity (0% to 100%) and specificity (31% to 100%). For anti-TB IgG, the only test with enough studies for meta-analysis, pooled sensitivity was 76% (95% CI 63%–87%) in smear-positive (seven studies) and 59% (95% CI 10%–96%) in smear-negative (four studies) patients; pooled specificities were 92% (95% CI 74%–98%) and 91% (95% CI 79%–96%), respectively. Compared with ELISA (pooled sensitivity 60% [95% CI 6%–65%]; pooled specificity 98% [95% CI 96%–99%]), immunochromatographic tests yielded lower pooled sensitivity (53%, 95% CI 42%–64%) and comparable pooled specificity (98%, 95% CI 94%–99%). For extrapulmonary TB, we included 25 studies (40% from low- and middle-income countries) with 1,809 participants. For all tests, estimates were variable for sensitivity (0% to 100%) and specificity (59% to 100%). Overall, quality of evidence was graded very low for studies of pulmonary and extrapulmonary TB.

Conclusions: Despite expansion of the literature since 2006, commercial serological tests continue to produce inconsistent and imprecise estimates of sensitivity and specificity. Quality of evidence remains very low. These data informed a recently published World Health Organization policy statement against serological tests.

Please see later in the article for the Editors' Summary.

Citation: Steingart KR, Flores LL, Dendukuri N, Schiller I, Laal S, et al. (2011) Commercial Serological Tests for the Diagnosis of Active Pulmonary and Extrapulmonary Tuberculosis: An Updated Systematic Review and Meta-Analysis. *PLoS Med* 8(8): e1001062. doi:10.1371/journal.pmed.1001062

Academic Editor: Carlton Evans, Universidad Peruana Cayetano Heredia, Peru

Received: November 30, 2010; **Accepted:** June 9, 2011; **Published:** August 9, 2011

Copyright: © 2011 World Health Organization; licensee Public Library of Science (PLoS). This is an Open Access article in the spirit of the Public Library of Science (PLoS) principles for Open Access <http://www.plos.org/oa/>, without any waiver of WHO's privileges and immunities under international law, convention, or agreement. This article should not be reproduced for use in association with the promotion of commercial products, services, or any legal entity. There should be no suggestion that WHO endorses any specific organization or products. The use of the WHO logo is not permitted. This notice should be preserved along with the article's original URL.

Funding: This systematic review was commissioned by the WHO with funding provided by USAID through a grant administered by the UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR). No funding bodies had any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have no financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed. KRS serves as Coordinator of the Evidence Synthesis and Policy subgroup of Stop TB Partnership's New Diagnostics Working Group (NDWG). AR served as the secretary of the NDWG (2006–10) and is employed by WHO/TDR, the agency that administered the grant used to fund this systematic review. AR contributed to the conception and design of the systematic review and critical revision of and decision to publish the manuscript. MP is a recipient of a New Investigator Award from the Canadian Institutes of Health Research (CIHR). This funding source had no role in the preparation of this manuscript, nor the decision to submit the manuscript for publication. MP serves as an external consultant for the Bill & Melinda Gates Foundation. MP also serves as a co-chair of the NDWG. KRS, AR, MP, and SL were involved in the WHO Expert Group process that reviewed the evidence on serological tests. KRS was involved as a systematic reviewer in the World Health Organization Expert Group process that reviewed the evidence on serological tests in July 2010. MP serves on the editorial boards of *PLoS Medicine* and *PLoS One*.

Abbreviations: ELISA, enzyme-linked immunosorbent assay; HSROC, hierarchical summary receiver operating characteristic; QUADAS, Quality Assessment of Diagnostic Accuracy Studies; TB, tuberculosis; WHO, World Health Organization

* E-mail: madhukar.pai@mcgill.ca

Introduction

Despite impressive advances in tuberculosis (TB) control over the last decade [1], missed diagnoses continue to fuel the global epidemic, leading to more severe illness for patients and enabling further transmission of *Mycobacterium tuberculosis* [2]. Smear microscopy and chest radiography, the primary tools used in resource-limited countries for identifying TB, often perform poorly, especially in HIV-coinfected patients [3–5]. Improved techniques, such as liquid culture for *M. tuberculosis* and nucleic acid amplification tests, are often too expensive and complex for routine use in resource-limited settings. The Xpert MTB/RIF (Cepheid), a new technology recently endorsed by the World Health Organization (WHO), provides high sensitivity for detection of TB and drug resistance [6]. WHO has issued a blueprint for Xpert's implementation [7]; however, high cost may be a barrier for scaling up this technology in many areas where the epidemic is most severe [2].

Serological tests have a long history and have been used successfully for the rapid diagnosis of many infectious diseases (e.g., HIV, syphilis, and viral hepatitis). In this paper, “serological tests” refers to blood tests that detect the humoral immune (antibody) responses to *M. tuberculosis* antigens. Serological tests are not to be confused with interferon-gamma release assays that measure the T-cell-based interferon-gamma response to *M. tuberculosis* antigens. In comparison with microscopy, serological tests appear to offer several advantages: (1) the result from a serological test using the enzyme-linked immunosorbent assay (ELISA) format could be available within hours, and the result using an immunochromatographic assay format, within minutes; (2) a serological test, if developed into a point-of-care test, could potentially replace microscopy or extend testing to lower levels of health services; and (3) in children, for whom sputum is difficult to obtain, and in patients suspected of having extrapulmonary TB, a blood test may be more practical.

Although currently the International Standards for TB Care discourages the use of serological tests in routine practice [8] and no international guideline recommends their use, dozens of commercial serological tests for TB diagnosis are offered for sale in many parts of the world [9], including Afghanistan, Bangladesh, Brazil, Cambodia, China, India, Indonesia, Kenya, Myanmar, Nigeria, Pakistan, Philippines, Russia, South Africa, Thailand, Uganda, and Viet Nam, as was recently found in a survey of 22 high TB burden countries [10]. For example, in India, numerous products with claims of high accuracy in their package inserts are available for purchase (Table S1), and an estimated 1.5 million serological tests are performed every year [10].

We are aware of four systematic reviews and one laboratory-based evaluation on this topic. The first review included only studies with a cohort or case series design and searched the literature through 2003 [11]. Performance of the tests was modest, and sensitivity decreased when only studies meeting at least two design-related criteria were included (seven studies, pooled sensitivity of 34%) [11]. Two subsequent reviews evaluating commercial serological tests for pulmonary TB (68 studies) [12] and extrapulmonary TB (21 studies) [13] found the sensitivity and specificity of these tests to be highly variable. The fourth review, a meta-analysis of in-house serological tests for the diagnosis of pulmonary TB (254 studies including 51 distinct single antigens and 30 distinct multiple-antigen combinations), identified potential candidate antigens for inclusion in an antibody-detection-based TB test in patients with and without HIV infection; however, no single antigen achieved sufficient sensitivity to replace smear microscopy [14]. A laboratory-based evaluation of 19 rapid commercial tests conducted by the

WHO Special Programme for Research and Training in Tropical Diseases found that, in comparison with culture plus clinical follow-up, serological tests provided low and variable sensitivity (1% to 60%) and specificity (53% to 99%) [15].

Since the publication of the previous reviews, the evidence base has grown and approaches to meta-analysis of diagnostic tests have evolved. This updated systematic review was commissioned by WHO to guide policy recommendations on serological tests for TB, with a special focus on the relevance of these assays in low- and middle-income countries. The objective of this review is to synthesize new evidence since 2006 in order to address the following question: what is the diagnostic accuracy of commercial serological tests for active TB (pulmonary and extrapulmonary TB) in adults and children, with and without HIV infection? Specifically, we were interested in evaluating the use of a serological assay as a replacement test for, or an additional test after, smear microscopy.

Methods

We followed methods for conducting and reporting systematic reviews and meta-analyses recommended by the Cochrane Collaboration Diagnostic Test Accuracy Working Group and the PRISMA statement (Text S1), including the preparation of a protocol and analysis plan (Text S2) [16–18].

Selection Criteria and Definitions

Types of studies. Diagnostic studies (with any study design) were included that evaluated serological tests for active TB (pulmonary and extrapulmonary TB) in patients who provided sera before or within 14 d of starting antituberculous treatment.

Participants. The participants constituted adults and children, with and without HIV infection, with suspected or confirmed active TB, from all clinical settings (clinic or hospital). The protocol for the current review included studies with at least ten TB cases. Studies could be performed in any country regardless of TB incidence or income status.

Index test. The index test was any commercial serological test for the diagnosis of active TB.

Comparator tests. There was either no test or smear microscopy used for comparison.

Target conditions. The target conditions were pulmonary and extrapulmonary TB.

Reference standards. Pulmonary TB required positivity on mycobacterial culture. (The previous review accepted positivity on either culture or smear microscopy as the reference standard [12].) Extrapulmonary TB required positivity on at least one of the following tests: culture, smear, or histopathological examination.

Outcomes. The outcomes were sensitivity and specificity. Sensitivity refers to the proportion of patients with a positive serological test result among patients with TB confirmed by the reference standard. Specificity refers to the proportion of participants with a negative serological test result among participants without TB according to the reference standard. To estimate specificity, we selected only one non-TB group if a study had more than one such group. The preferred non-TB participants were those in whom active TB was initially suspected but later ruled out (“other respiratory disease” or “mixed disease” groups), and who were from the same population as TB patients.

Extrapulmonary TB. Extrapulmonary TB was classified as lymph node, pleural, meningeal and/or central nervous system, bone and/or joint, genitourinary, abdominal, skin, other sites,

disseminated, and multiple sites (extrapulmonary TB cases from different sites are combined to obtain at least ten extrapulmonary TB cases).

Country income status. Country income status was classified according to the World Bank List of Economies [19].

Exclusion criteria. The following studies were excluded: (1) studies published before 1990; (2) animal studies; (3) conference abstracts and proceedings; (4) studies on the detection of latent TB infection; (5) studies on nontuberculous mycobacterial infection; (6) studies that used non-immunological methods for detection of antibodies; and (7) basic science literature that focused on detection/cloning of new antigens or their immunological properties (i.e., early pre-clinical studies).

Search Methods

We updated the database searches (MEDLINE [1 May 2006 to 29 June 2010], BIOSIS [1 January 2005 to 10 February 2010], EMBASE [1 October 2005 to 10 February 2010], and Web of Science [1 January 2005 to 10 February 2010]) that were carried out in previous systematic reviews (MEDLINE [1 January 1990 to 30 May 2006], BIOSIS [1 January 1990 to 6 December 2005], EMBASE [1 January 1990 to 11 October 2005], and Web of Science [1 January 1990 to 6 December 2005]) for relevant studies that reported data on commercial serological tests for active TB. The original search was limited to English, and the updated search was performed without a language restriction. The search field tags used in database searching were MeSH terms (mh), title/abstract words (tiab), and title (ti). The terms used included: tuberculosis[mh] OR mycobacterium tuberculosis[mh], “sensitivity and specificity”[mh] OR diagnostic*[tiab] OR predictive value of tests[mh] OR immunologic tests[mh] OR immunochemistry[mh] OR serology[ti] OR serological[ti] OR serodiagnosis[tiab] OR serodiagnostic[tiab] OR immunodiagnosis[tiab] OR immunodiagnostic[tiab] OR antibody[tiab] OR antibodies[tiab] OR elisa[tiab] OR immunosorbent[tiab] OR (western[tiab] AND blot*[tiab]) OR immunoassay[tiab] OR “humoral immune” OR “humoral immunity” OR “humoral antibody” OR “immune based” OR “antibody detection” (Text S3).

In addition to database searches, we also searched reference lists of eligible papers and related reviews, and contacted authors and researchers in the field to identify additional potentially relevant published studies. For lack of time, we did not specifically seek to identify unpublished studies.

Study Selection

Initially, two reviewers (KRS and LLF) independently screened the accumulated citations for relevance and then independently reviewed full-text articles using prespecified eligibility criteria. Disagreements about study selection were resolved by discussion.

Data Extraction

A data extraction form was created and pilot-tested with a subset of eligible studies and then finalized. Two reviewers independently extracted data from included studies with the standardized form on the following characteristics: study design; age group (children <15 y of age); HIV status; case country of residence; sputum smear status (pulmonary TB); site of TB (extrapulmonary TB); assay type (e.g., ELISA, immunochromatographic test); antibody class detected (IgG, IgM, and IgA); serological test name; antigen composition; condition of the specimen (fresh or frozen); and sensitivity and specificity (data were extracted as true positives, false positives, false negatives, and true negatives). In some cases, study investigators evaluated more than one diagnostic test with the same set of participants. In these situations, we extracted data for each test and

considered each dataset to be an independent study. For example, Anderson et al. [20] contributed three studies evaluating three serological tests: (1) InBios Active TbDetect IgG ELISA (InBios International); (2) IBL *M. tuberculosis* IgG ELISA (IBL-Hamburg), and (3) anda-TB IgG (Anda Biologicals) [20]. The agreement between reviewers on data extraction for sensitivity and specificity was 100%. Other differences between the reviewers (these differences mainly concerned the methodological quality assessment) were resolved by discussion. When necessary, we contacted authors of papers identified through the updated literature search for additional information.

While extracting data, we looked for studies that considered the added value of serological tests to determine if they contributed to active TB diagnosis beyond that ascertained by conventional tests such as symptoms, sputum smears, and chest radiographs. In particular, we looked for studies comparing microscopy with microscopy plus serology or studies that performed multivariable analysis. Since we did not identify any studies of this type, we considered studies in smear-negative patients to provide indirect evidence of the use of serology as an add-on test to microscopy. In addition, we looked for information on patient-important outcomes. Patient-important outcomes for this review could include an increased number of TB patients detected, decreased time to starting treatment, increased number of patients starting TB treatment, decreased number of false-positive TB patients treated, and decreased number of patients lost because of a reduced number of visits. Finally, we looked for information on the values and preferences of patients associated with these tests.

Assessment of the Methodological Quality of Individual Studies

Two reviewers (KRS and LLF) independently assessed study quality using the core set of 11 items from Quality Assessment of Diagnostic Accuracy Studies (QUADAS), a validated tool to evaluate the presence of bias and variation in diagnostic accuracy studies [21]. As recommended, we scored each item as “yes,” “no,” or “unclear.” We considered representative patient spectrum (i.e., was the spectrum of patients representative of the patients who will receive the test in practice?) to be persons suspected of having active TB who were consecutively or randomly enrolled. For pulmonary TB, a score of “yes” for representative spectrum also required that patients were evaluated in an outpatient setting. For all studies, we scored the following six items as “yes”: acceptable reference standard (the reference standard was a criterion for inclusion); acceptable delay between serological test and reference standard; partial verification avoided; incorporation avoided; reference standard results blinded (as culture result was considered to be entirely objective in interpretation); and relevant clinical information. “Differential verification avoided” was scored as “yes” if all participants suspected of having TB were evaluated with the same reference standard or if participants without TB were reported to be asymptomatic and healthy. “Index test (serological test) result blinded” (to reference standard result) was scored as “yes” if this was explicitly stated in the paper. “Uninterpretable results reported” was scored as “yes” if uninterpretable results were described or there was a statement about the absence of uninterpretable results. “Withdrawals explained” was scored as “yes” if a flow diagram or statement was included making it clear what happened to all participants in the study. Conflicts of interest are known to be a concern in diagnostic studies [22]. Therefore, we evaluated the involvement of test manufacturers. Finally, we grouped studies according to the type of serological test or site of

TB (in the case of extrapulmonary TB) and assessed study quality separately for each subgroup.

GRADE Quality of Evidence

We used the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach, a transparent and systematic process for making judgments about quality of the evidence [23]. GRADE specifies four categories for quality: high, moderate, low, and very low. These categories are then applied to the body of evidence for each outcome, rather than to individual studies. In the GRADE approach, the categories reflect the extent of confidence that an estimate of effect is correct [24]. Quality begins with a consideration of study design (e.g., randomized controlled trials and cross-sectional studies in patients with diagnostic uncertainty are considered high quality) and may be compromised by five factors: limitations (risk of bias assessed by QUADAS, such as absence of consecutive or random selection of participants and lack of blinding of test results), indirectness (lack of generalizability and use of test results as surrogates for patient-important outcomes), inconsistency (unexplained heterogeneity), imprecision (wide confidence intervals for estimates of test accuracy), and risk of publication bias [23].

Data Analysis

Descriptive analyses were performed using SPSS (version 14.0.1.366). For each study, the sensitivity and specificity of the serological test along with the 95% CIs were calculated, and forest plots were generated to display sensitivity and specificity estimates using Review Manager 5.0 (The Nordic Cochrane Center). Heterogeneity was assessed visually using forest plots (Review Manager 5.0).

Selection of Subgroups for Meta-Analysis

We recognized that studies were heterogeneous in many respects, particularly concerning the serological test used, antibody class detected, sputum smear status (pulmonary TB), and site of extrapulmonary TB. Therefore, in order to address heterogeneity and combine study results, subgroups of “comparable” tests and extrapulmonary sites were prespecified. When possible, studies were stratified by smear and HIV status. For meta-analysis, at least four studies were required to be available for inclusion in a subgroup, in order to strengthen results and reduce the possibility of finding a significant result by chance. This classification resulted in seven subgroups for meta-analysis (four pulmonary and three extrapulmonary TB subgroups). As noted above, in some cases, study investigators evaluated more than one diagnostic test with the same participants; therefore, some meta-analyses included the same individuals multiple times.

To summarize test performance within each subgroup, we carried out bivariate meta-analyses that jointly modeled sensitivity and specificity. These models weighted studies according to the sampling variability within studies as well as the unexplained heterogeneity between studies using a random effects approach [25]. Subgroups were considered homogeneous with respect to a number of observed variables. However, within a subgroup, it is likely that the heterogeneity between studies could be explained by measurable but unobserved quantities, e.g., the positivity cutoff, that we could not address. Therefore, for the pooled results of studies in the meta-analysis, we did not attempt to quantify unobserved heterogeneity using statistics such as I^2 or chi-squared [18]. The model was estimated using a Bayesian approach with nonsubjective prior distributions and implemented using WinBUGS (version 1.4.1) [26]. We used Wilson’s method for

estimating the credible interval as this method performs well even when the probability or the sample size is small [27]. Finally, a hierarchical summary receiver operating characteristic (HSROC) curve was plotted for selected meta-analyses. The HSROC curve plots sensitivity versus specificity and provides information on the overall performance of a test across different thresholds. The closer the curve is to the upper left-hand corner of the plot (sensitivity and specificity are both 100%), the better the performance of the test [28]. The plots were made using R (version 2.6.1) [29].

Results

Pulmonary TB

Results of the search. Initially, 4,256 citations were identified (Figure 1). After screening titles and abstracts, 160 potentially relevant full-text papers were retrieved. Thirty-one papers (20 from the original review and 11 from the update) describing 67 studies and involving 5,147 participants (sample size = 8,318) were included in the review [20,30–59]. A list of excluded articles with their reasons for exclusion is provided in Text S4.

Included studies. Of 67 total studies, six (9%) were reported in languages other than English: Spanish (2), Turkish (1), Chinese (1), Bosnian (1), and Russian (1). Thirty-two (48%) studies were conducted in low- and middle-income countries. No studies were randomized controlled trials; 55% of studies used a cross-sectional study design, and 45% of studies used a case-control study design. All but one study reported recruiting TB and non-TB patients from the same underlying population [56]. One study involved HIV-infected individuals, and no studies involved children. Thirty-one (46%) studies involved smear-positive patients, 28 (42%) studies involved smear-negative patients, and eight (12%) studies involved patients with unspecified smear status. Fifty-four (81%) studies used ELISA, 12 (18%) studies used an immunochromatographic assay, and one study used a kaolin precipitation test. The majority of studies detected only IgG antibody (44 studies) and used frozen serum (51 studies). The median number of TB patients included in each study was 41 (interquartile range 33 to 54). Eighteen serological tests were included; anda-TB (IgG, IgA, and IgM) was the test most frequently evaluated (16/67 [24%]) (Table 1). The antigen composition for five (28%) of the total 18 tests was considered proprietary information. Of the tests with known antigens, all had unique antigenic compositions except for anda-TB and Hexagon, which both contained antigen A60.

One study directly compared a serological test to sputum microscopy. No studies evaluated the incremental value of adding a serological test after smear microscopy. However, as noted, 28 (42%) studies involved smear-negative patients. These studies were considered a proxy for a diagnostic strategy using serological tests in addition to microscopy. No studies reported on patient-important outcomes or patient values and preferences concerning these tests. Characteristics of included studies are described in Table S2.

Methodological quality, all included studies. As assessed with QUADAS, studies had very serious limitations. Of the total 67 studies only 19 (28%) were considered to include a representative patient population (we scored this item as “yes” when ambulatory patients suspected of having active TB were randomly or consecutively selected), and 34 (51%) studies reported blinding of the serological test result. The majority (60%) of studies reported industry involvement, mainly the donation of test kits (Figure 2). We downgraded two points for limitations in the GRADE Evidence Profile (Table 3).

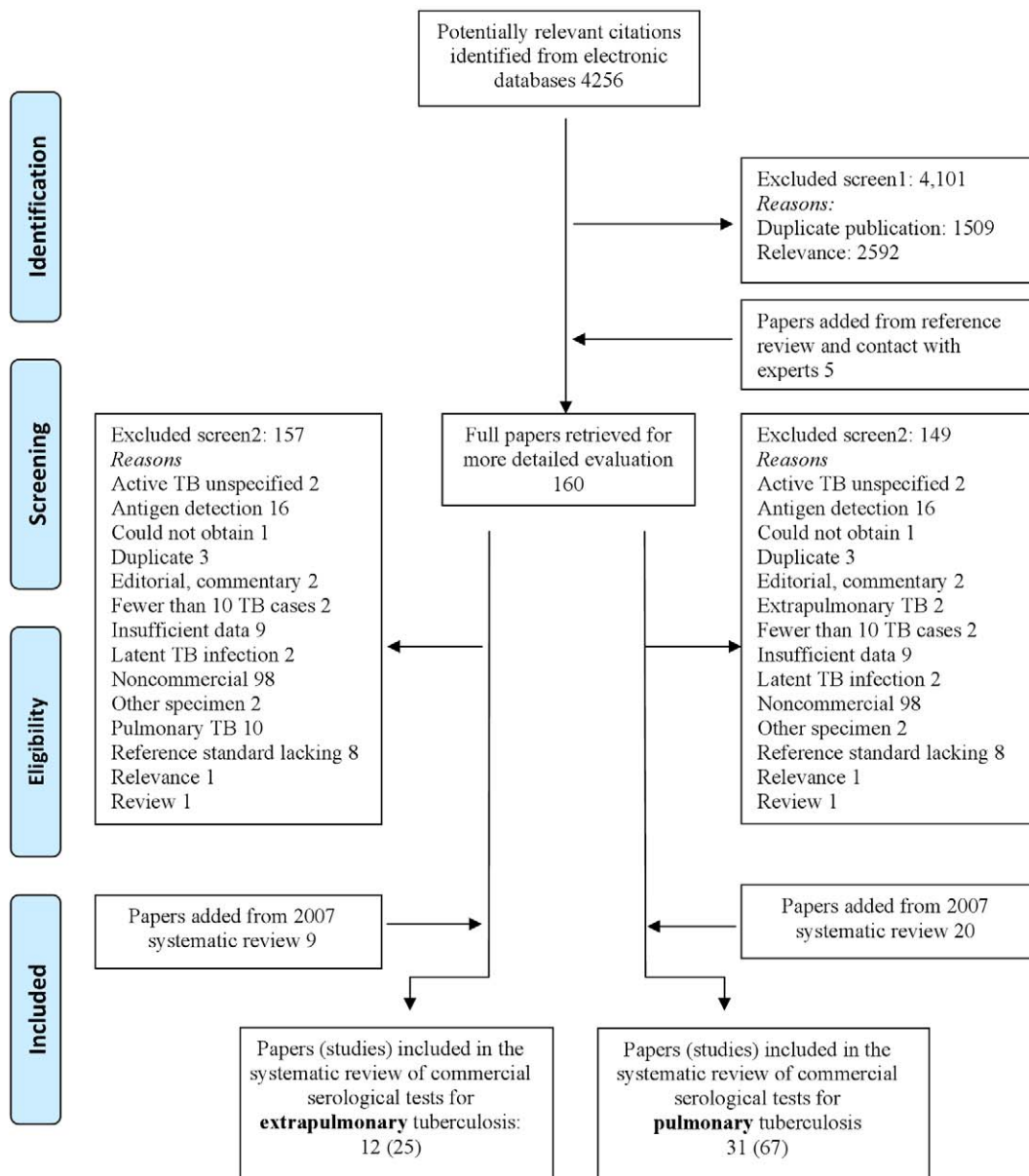


Figure 1. Flow of studies in the review of commercial serological tests for the diagnosis of active tuberculosis.
doi:10.1371/journal.pmed.1001062.g001

Test performance, all studies. As seen from the forest plots in Figure 3, studies displayed considerable heterogeneity, with sensitivity values ranging from 0% to 100% and specificity values, from 31% to 100%. We did not pool accuracy estimates because of the heterogeneity among studies. Similarly, when restricted to studies conducted in low- and middle-income countries, sensitivity (16% to 91%) and specificity (31% to 100%) were highly variable (data not shown).

Methodological quality, studies in smear-negative patients. As assessed with QUADAS, only 14 (50%) of the total 28 studies were considered to include a representative patient population. A majority (75%) of studies reported blinding of the serological test result (Figure 4). We downgraded one point for limitations in the GRADE Evidence Profile (Table 4).

Test performance, studies in smear-negative patients. For individual studies involving smear-negative patients, sensitivity values ranged from 29% to 77%, and specificity values, from 77% to 100% (Figure 5). We did not pool accuracy estimates because of the considerable heterogeneity among studies. As noted above, studies involving smear-negative patients were considered to provide indirect evidence of a diagnostic strategy using microscopy plus serology. Hence, as an add-on test, serological tests provided inconsistent sensitivity and specificity.

Analysis by subgroups. According to our prespecified analysis plan, there was a sufficient number of studies to perform a meta-analysis for only one serological test, and-a-TB IgG, with results stratified by smear status (seven studies of smear-positive and four studies of smear-negative patients). In studies of

Table 1. Commercial serological tests included in this systematic review for pulmonary TB.

Name of Test (Number of Studies)	Name of Manufacturer ^a , City, Country	Antigenic Composition	Antibodies Detected	Laboratory Technique
Detect-TB (2)	Adaltis—Advanced Laboratory Diagnostics Systems, Rome, Italy	Proprietary, a cocktail of three <i>M. tuberculosis</i> recombinant proteins and two synthetic peptides from five different proteins	IgG	ELISA
anda-TB ELISA ^b (16)	Anda Biologicals, Strasbourg, France	A60	IgG, IgM, IgA, IgG plus M	ELISA
Tuberculosis Specific Antigen (1)	Chengdu Pharmaceutical, Chengdu, China	Proprietary	Unknown	ELISA
Kaolin Agglutination Test (1)	Hitech Laboratories, Bombay, India	Tuberculophosphatide	IgG	Agglutination test
Hexagon TB (1)	Human Gesellschaft für Biochemica und Diagnostica, Wiesbaden, Germany	A60	IgG, IgA, IgM	ICT
Mycobacterium tuberculosis IgG (1)	IBL, Hamburg, Germany	18 kDa, 36 kDa, and 40 kDa recombinant antigens	IgG, IgA, IgM	ELISA
ICT TB (3)	ICT Diagnostics, Sydney, Australia	38 kDa and four proprietary antigens; all five antigens are recombinant	IgG	ICT
ActiveTBDetect (1)	InBios International, Seattle, US	Mtb81, Mtb8, Mtb48, DPEP (MPT32), 38 kDa protein, and two additional proprietary antigens	IgG	ELISA
SEVA (1)	Jamnala Bajaj Tropical Disease Research Centre, Mahatma Gandhi Institute of Medical Sciences, Maharashtra, India	31 kDa, native glycoprotein antigen from culture filtrate of MTB H37Rv	IgG	ELISA
TB Enzyme Immunoassay (5)	Kreatech, Amsterdam, The Netherlands	Kp-90 antigenic compound: LAM, 10 kDa, 16 kDa, 21 kDa, 30 kDa, 34 kDa, 65 kDa, and 95 kDa	IgG, IgA	ELISA
Determiner TB Glycolipid Assay (6)	Kyowa Medex, Tokyo, Japan	Contains trehalose 6,6'-dimycolate, trehalose monomycolate, diacyltrehalose, phenolic glycolipid, 2,3,6,6-tetraacyl-trehalose-2-sulfate, and 2,3,6-triacyl-trehalose	IgG	ELISA
Assure TB (2)	MedTek, Parangue City, Philippines; and Genelabs Diagnostics, Singapore	Proprietary, two recombinant antigens	IgG	ICT
MycDot (3)	Mossman Associates, Milford, Massachusetts, US	LAM	IgG	ICT
Pathozyme Myco ^c (11)	Omega Diagnostics, Alva, Scotland	LAM, recombinant 38 kDa	IgG, IgA, IgM	ELISA
Pathozyme TB Complex (2)	Omega Diagnostics, Alva, Scotland	Recombinant 38 kDa	IgG	ELISA
Pathozyme TB Complex Plus ^d (9)	Omega Diagnostics, Alva, Scotland	Recombinant 38 kDa and 16 kDa	IgG	ELISA
SDHO MTB (1)	SDHO Laboratories, Saint-Sauveur des Monts, Canada	Proprietary	IgG	ICT
Serocheck-MTB (1)	Zephyr Biomedicals, Verna, India	Recombinant 14 kDa, 38 kDa, 16 kDa, and 6 kDa	Unknown	ICT

ICT, immunochromatographic test; LAM, lipoarabinomannan.

^aSome manufacturers may no longer provide the serological tests listed above. It is also possible that the same test may be marketed under various names by different companies.

^banda-TB: IgG (13 studies), IgM (one study), IgA (two studies).

^cPathozyme Myco: IgG (three studies), IgM (two studies), IgA (two studies), IgG plus IgM (one study), IgG plus IgA (one study), IgM plus IgA (one study), IgG plus IgM plus IgA (one study).

^dPathozyme TB Complex Plus as a single test (three studies) or in combination with Pathozyme Myco G (one study), Pathozyme Myco M (one study), Pathozyme Myco A (one study), Pathozyme Myco G and A (one study), Pathozyme Myco M and A (one study), Pathozyme Myco G, M, and A (one study).

doi:10.1371/journal.pmed.1001062.t001

smear-positive patients, one study was conducted in a low-income country [43]. In studies of smear-negative patients, no studies were conducted in a low- or middle-income country.

Methodological quality of studies evaluating anda-TB IgG. In studies of smear-positive patients, no studies were considered to have a representative patient population (participants were known TB cases rather than suspected cases, were inpatients, and/or were enrolled by convenience), and only two studies reported blinding of the serological test result [43,52] (Figure S1). In studies of smear-negative patients, no studies were considered to have a representative patient population (participant selection was by

convenience or not reported), and only one study reported blinding of the serological test result [52] (Figure S2).

Meta-analysis. In studies involving smear-positive patients, anda-TB IgG yielded a pooled sensitivity of 76% (95% CI 63–87) and a pooled specificity of 92% (95% CI 74–98). In studies involving smear-negative patients, the pooled sensitivity of anda-TB IgG decreased to 59% (95% CI 10–96); the 95% CI was very wide, reflecting the imprecision of the sensitivity estimate. Pooled specificity was 91% (95% CI 79–96) (Table 2). The HSROC curves show the decreased performance of the test in smear-negative patients compared with smear-positive patients (Figure 6).

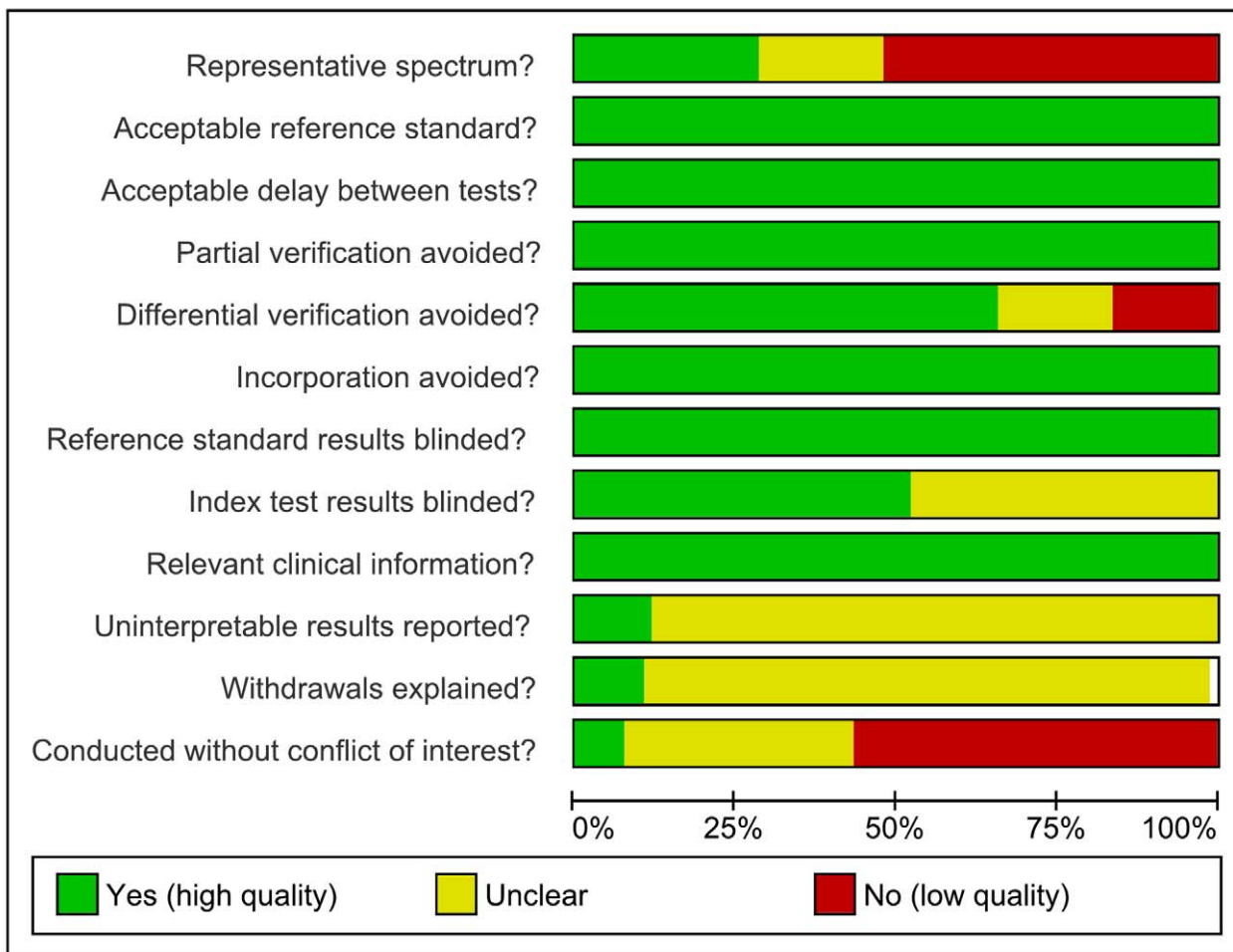


Figure 2. Methodological quality graph, all studies, pulmonary TB. Review authors’ judgments about each methodological quality item. doi:10.1371/journal.pmed.1001062.g002

We explored whether test technique had an impact on accuracy. Compared with ELISA (pooled sensitivity 60% [95% CI 6–65]; pooled specificity 98% [95% CI 96–99]), immunochromatographic assays yielded lower pooled sensitivity (53%, 95% CI 42–64) and comparable pooled specificity (98%, 95% CI 94–99)

(Table 2; Figure 7). The probability that the pooled sensitivity of ELISA tests exceeds that of immunochromatographic assays was estimated at 0.88.

TB in HIV-infected patients. The only study identified involving HIV-infected patients compared the performance of the

Table 2. Bivariate meta-analyses: pooled sensitivity and specificity estimates by subgroup.

Subgroup	Number of Studies	Number of Participants (Sample Size)	Sensitivity	Specificity
Pulmonary TB				
anda-TB IgG, smear-positive	7	870 (870)	76 (63–87)	92 (74–98)
anda-TB IgG, smear-negative	4	700 (700)	59 (10–96)	91 (79–96)
ELISA	54	3,696 (6,434)	60 (6–65)	98 (96–99)
Immunochromatographic Test ^a	12	1,231 (1,512)	53 (42–64)	98 (94–99)
Extrapulmonary TB				
Lymph node TB	6	640 (922)	64 (28–92)	90 (76–97)
Pleural TB	5	322 (572)	46 (29–63)	87 (51–99)
anda-TB IgG	10	1,055 (1,637)	81 (49–97)	85 (77–92)

Sensitivity and specificity estimates given as posterior means (percent) with 95% credible intervals in parentheses.

^aSerological tests included: ICT TB (three studies), Assure TB (two studies), MycoDot (three studies), SDHO (two studies), Hexagon (one study), Serocheck-MTB (one study).

doi:10.1371/journal.pmed.1001062.t002

Table 3. GRADE evidence profile: should commercial serological tests be used as a replacement test for conventional tests such as smear microscopy in patients of any age suspected of having pulmonary tuberculosis?

Outcome	Number of Studies (Participants)	Study Design	Limitations	Indirectness	Inconsistency	Imprecision	Publication Bias	Final Quality	Effect per 1,000 ^a	Importance ^b
True Positives	67 (5,147)	Cross-sectional and case-control	Very serious ^c (-2)	No serious indirectness ^d	Very serious ^e (-2)	Serious ^f	Likely ^g	Very low ⊕○○○	Prevalence 10%: 64; prevalence 30%: 192	Critical
True Negatives	67 (5,147)	Cross-sectional and case-control	Very serious ^c (-2)	No serious indirectness ^d	Very serious ^e (-2)	Serious ^f	Likely ^g	Very low ⊕○○○	Prevalence 10%: 819; prevalence 30%: 637	Critical
False Positives	67 (5,147)	Cross-sectional and case-control	Very serious ^c (-2)	No serious indirectness ^d	Very serious ^e (-2)	Serious ^f	Likely ^g	Very low ⊕○○○	Prevalence 10%: 81; prevalence 30%: 63	Critical
False Negatives	67 (5,147)	Cross-sectional and case-control	Very serious ^c (-2)	No serious indirectness ^d	Very serious ^e (-2)	Serious ^f	Likely ^g	Very low ⊕○○○	Prevalence 10%: 36; prevalence 30%: 108	Critical

Based on sample size = 8,318, sensitivity median = 64%, specificity median = 91%. The quality of evidence was rated as high (no points subtracted), moderate (one point subtracted), low (two points subtracted), or very low (>2 points subtracted) based on five factors: study limitations, indirectness of evidence, inconsistency in results across studies, imprecision in summary estimates, and likelihood of publication bias. For each outcome, the quality of evidence started at high when there were randomized controlled trials or high-quality observational studies (cross-sectional or cohort studies enrolling patients with diagnostic uncertainty) and at moderate when these types of studies were absent. No points were subtracted when there were negligible issues identified; one point was subtracted when there was a serious issue identified; two points were subtracted when there was a very serious issue identified in any of the criteria used to judge the quality of evidence. Points subtracted are in parentheses. Publication bias was rated as "not likely," "likely," or "very likely" [23].

^aWhat do these results mean given 10% or 30% disease prevalence among individuals being screened for TB?

^bOutcomes were ranked by their relative importance as critical, important, or of limited importance. Ranking helped to focus attention on those outcomes that were considered most important.

^cThe majority of studies lacked a representative patient population and were not blinded.

^dAlthough diagnostic accuracy is considered a surrogate for patient-important outcomes, we did not downgrade.

^eThere was considerable heterogeneity in study results.

^fWe did not pool accuracy estimates. The 95% CIs were wide for many individual studies. We did not downgrade as there were a large number of studies and we had already taken off two points for inconsistency.

^gData included in the review did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests. Therefore, publication bias cannot be ruled out. It is prudent to assume some degree of publication bias as studies showing poor performance of serological tests were probably less likely to be published. No points were deducted.

doi:10.1371/journal.pmed.1001062.t003

SDHO MTB test (SDHO Laboratories) head-to-head with smear microscopy in 55 individuals suspected of having pulmonary TB residing in the Central African Republic [44]. TB was confirmed by culture. Compared with smear microscopy (sensitivity 68% [95% CI 49–83]), SDHO yielded a sensitivity of only 16% (95% CI 5–34). Specificity of SDHO was 90% (95% CI 74–98), lower than the specificity of smear microscopy (100% [95% CI 89–100]).

Extrapulmonary TB

Results of the search. Twelve papers (nine from the original review and three from the update) describing 25 studies and involving 1,809 participants (sample size = 3,776) were included in the review (Figure 1) [45,48,60–69].

Included studies. Of 25 total studies, ten (40%) studies were conducted in low- and middle-income countries. All papers were written in English. Only one study involved HIV-infected individuals. The vast majority (88%) of studies involved adults. In two studies (reported from one paper), 13 of 35 extrapulmonary TB cases occurred in children; however, data were not provided separately for the children [67]. One study specified 13 y as the minimum age for eligibility; however, the age range for the enrolled participants was not reported [64]. Of 25 total studies, serological tests were evaluated for diagnosis of the following forms of extrapulmonary TB: lymph node, six studies; pleural, five studies; multiple sites, five studies (see Table S3 for a list of sites

involved); genitourinary, two studies; disseminated, four studies; and meningeal, one study. In two studies, the site of extrapulmonary involvement was not reported. Six distinct serological tests were evaluated; 17 (68%) of the total 25 studies used anti-TB (IgG, ten studies; IgM, five studies; IgA, one study; IgM plus IgA, one study). ELISA was used in 21 (84%) studies, and immunochromatographic assays, in four studies. The majority (72%) of studies detected IgG antibodies. The condition of the specimen was frozen in six (24%) studies and not reported in 19 studies. The median number of TB patients included in each study was 35 (interquartile range 30 to 56). No studies reported on patient-important outcomes or patient values and preferences concerning these tests. Characteristics of included studies are described in Table S3.

Methodological quality, all included studies. Of the 25 total studies, only one (4%) study was considered to include a representative patient population, and only four (16%) studies reported blinding of the serological test result (Figure 8).

Test performance, all studies. As seen from the forest plots in Figure 9, studies displayed considerable heterogeneity, with sensitivity values ranging from 0% to 100%, and specificity values, from 59% to 100%. We did not pool accuracy estimates because of the heterogeneity among studies.

Analysis by subgroups. There was a sufficient number of studies to perform a meta-analysis for only one serological test,

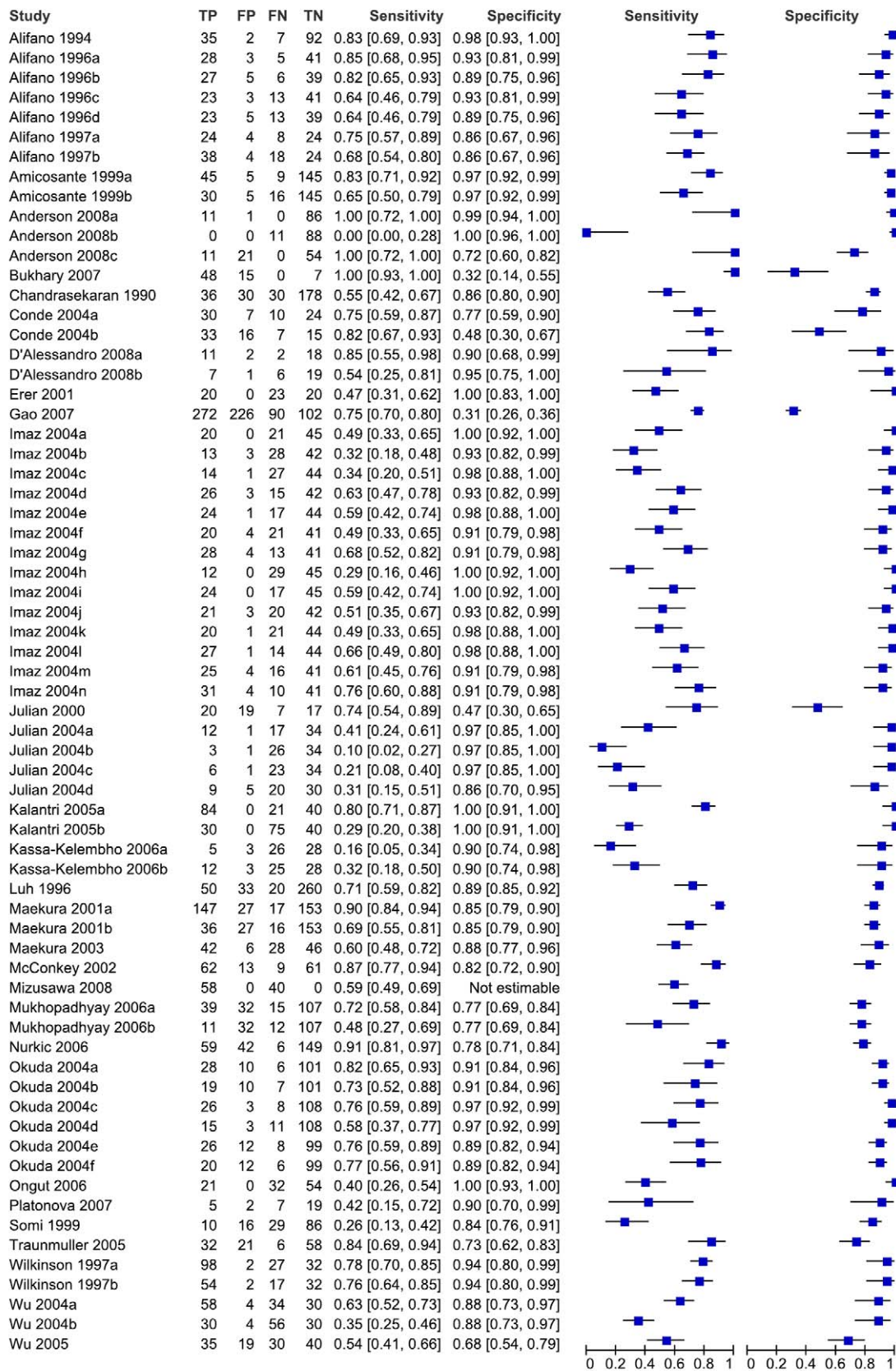


Figure 3. Forest plots of sensitivity and specificity, all studies, pulmonary TB. Specificity data for Mizusawa et al. [49] were not reported. FN, false negative; FP, false positive; TN, true negative; TP, true positive. 95% CIs are included in brackets. On the right, the sensitivity and specificity estimates for individual studies are indicated by blue squares and the 95% CIs are indicated by black horizontal lines.
doi:10.1371/journal.pmed.1001062.g003

	Representative spectrum?	Acceptable reference standard?	Acceptable delay between tests?	Partial verification avoided?	Differential verification avoided?	Incorporation avoided?	Reference standard results blinded?	Index test results blinded?	Relevant clinical information?	Uninterpretable results reported?	Withdrawals explained?	Conducted without conflict of interest?
Alifano 1996c	+	+	+	+	?	+	+	?	+	?	?	?
Alifano 1996d	+	+	+	+	?	+	+	?	+	?	?	?
Alifano 1997b	+	+	+	+	-	+	+	?	+	?	?	?
Amicosante 1999b	+	+	+	+	?	+	+	?	+	?	?	?
Chandrasekaran 1990	?	+	+	+	+	+	+	?	+	+	?	-
Imaz 2004a	+	+	+	+	+	+	+	+	+	?	?	-
Imaz 2004b	+	+	+	+	+	+	+	+	+	?	?	-
Imaz 2004c	+	+	+	+	+	+	+	+	+	?	?	-
Imaz 2004d	+	+	+	+	+	+	+	+	+	?	?	-
Imaz 2004e	+	+	+	+	+	+	+	+	+	?	?	-
Imaz 2004f	+	+	+	+	+	+	+	+	+	?	?	-
Imaz 2004g	+	+	+	+	+	+	+	+	+	?	?	-
Imaz 2004h	+	+	+	+	+	+	+	+	+	?	?	-
Imaz 2004i	+	+	+	+	+	+	+	+	+	?	?	-
Imaz 2004j	+	+	+	+	+	+	+	+	+	?	?	-
Imaz 2004k	+	+	+	+	+	+	+	+	+	?	?	-
Imaz 2004l	+	+	+	+	+	+	+	+	+	?	?	-
Imaz 2004m	+	+	+	+	+	+	+	+	+	?	?	-
Imaz 2004n	+	+	+	+	+	+	+	+	+	?	?	-
Luh 1996	-	+	+	+	?	+	+	?	+	?	?	?
Maekura 2001b	-	+	+	+	-	+	+	+	+	?	?	-
Maekura 2003	-	+	+	+	+	+	+	+	+	?	?	-
Mukhopadhyay 2006b	-	+	+	+	+	+	+	+	+	?	+	-
Okuda 2004b	?	+	+	+	-	+	+	+	+	?	?	-
Okuda 2004d	?	+	+	+	-	+	+	+	+	?	?	-
Okuda 2004f	?	+	+	+	-	+	+	+	+	?	?	-
Wilkinson 1997b	-	+	+	+	+	+	+	+	+	?	?	-
Wu 2004b	?	+	+	+	+	+	+	?	+	?	?	?

Figure 4. Methodological quality summary, pulmonary TB, studies of smear-negative patients. Review authors' judgments about each methodological quality item. doi:10.1371/journal.pmed.1001062.g004

anda-TB IgG (ten studies). Eight studies were conducted in high-income countries, one study was conducted in India [61], and one study, in Turkey [66].

Methodological quality of studies evaluating anda-TB IgG. As assessed with QUADAS, studies had very serious limitations. No studies were considered to have a representative population (no studies reported selecting participants in a consecutive or random manner), and no studies reported interpreting the results of the serological test result without knowledge of the results of the reference standard.

Meta-analysis. We determined pooled sensitivity and specificity estimates for studies evaluating any serological test for the diagnosis of lymph node or pleural TB and for studies using anda-TB IgG (all extrapulmonary sites). For lymph node TB, pooled sensitivity was 64% (95% CI 28–92) and pooled specificity was 90% (95% CI 76–97). For pleural TB, pooled sensitivity was 46% (95% CI 29–63) and pooled specificity was 87% (95% CI 51–99). For anda-TB IgG, pooled sensitivity was 81% (95% CI 49–97) and pooled specificity was 85% (95% CI 77–92) (Table 2). For all three meta-analyses, the 95% CIs for sensitivity were wide, reflecting the imprecision of the estimates.

TB in HIV-infected patients. The only study identified involving HIV-infected patients evaluated the performance of the MycoDot test (Mossman Associates) in a cross-sectional study of patients suspected of having TB in Thailand [68]. In all, 142 HIV-infected (mean CD4 cell count = 188 cells/mm³ [range 7 to 632]) and 144 HIV-uninfected patients with newly diagnosed TB

participated in the study, of whom 50 patients (40 HIV-infected and ten HIV-uninfected patients) had a diagnosis of lymph node TB established by culture or histopathological examination. Compared with the sensitivity of MycoDot in HIV-uninfected TB patients (80%, 95% CI 44–98), the sensitivity of the test in HIV-infected TB patients was considerably lower (33%, 95% CI 19–39). The specificity in both groups was 97% (95% CI 93–99) (Table S3).

GRADE Evidence Profiles

For the pulmonary TB studies, the quality of the body of evidence supporting TB serology’s estimates of sensitivity and specificity was graded as “very low” (Tables 3 and 4). Thus, regardless of the width of the 95% CIs (which reflects the size of studies and the standard deviation of their measured results), we have very low confidence in the estimates obtained from pooling studies in the meta-analysis [23]. For the extrapulmonary TB studies, the final quality grades were also very low (data not shown).

Discussion

This updated systematic review assessing the diagnostic accuracy of commercial serological tests for pulmonary and extrapulmonary TB summarizes the current literature and includes 14 new papers (approximately 30% of the included papers) identified since our previous reviews [12,13]. Unlike the earlier reviews, in the update, we performed a meta-analysis using a bivariate random effects model to account for the variability in

Table 4. GRADE evidence profile: should commercial serological tests be used as an “add on” test to smear microscopy in patients of any age suspected of having pulmonary TB?

Outcome	Number of Studies (Participants)	Study Design	Limitations	Indirectness	Inconsistency	Imprecision	Publication Bias	Final Quality	Effect per 1,000 ^a	Importance ^b
True Positives	28 (1,961)	Mainly cross-sectional	Serious ^c (-1)	Serious ^d (-1)	Very serious ^e (-2)	Serious ^f	Likely ^g	Very Low ⊕○○○	Prevalence 10%: 61	Critical
True Negatives	28 (1,961)	Mainly cross-sectional	Serious ^c (-1)	Serious ^d (-1)	Very serious ^e (-2)	Serious ^f	Likely ^g	Very Low ⊕○○○	Prevalence 10%: 828	Critical
False Positives	28 (1,961)	Mainly cross-sectional	Serious ^c (-1)	Serious ^d (-1)	Very serious ^e (-2)	Serious ^f	Likely ^g	Very Low ⊕○○○	Prevalence 10%: 72	Critical
False Negatives	28 (1,961)	Mainly cross-sectional	Serious ^c (-1)	Serious ^d (-1)	Very serious ^e (-2)	Serious ^f	Likely ^g	Very Low ⊕○○○	Prevalence 10%: 39	Critical

This table includes studies conducted in smear-negative patients as a proxy for a diagnostic strategy using serological tests in addition to smear microscopy. Based on sample size = 3,433, sensitivity median = 61% and specificity median = 92%. The quality of evidence was rated as high (no points subtracted), moderate (one point subtracted), low (two points subtracted), or very low (>2 points subtracted) based on five factors: study limitations, indirectness of evidence, inconsistency in results across studies, imprecision in summary estimates, and likelihood of publication bias. For each outcome, the quality of evidence started at high when there were randomized controlled trials or high-quality observational studies (cross-sectional or cohort studies enrolling patients with diagnostic uncertainty) and at moderate when these types of studies were absent. No points were subtracted when there were negligible issues identified; one point was subtracted when there was a serious issue identified; two points were subtracted when there was a very serious issue identified in any of the criteria used to judge the quality of evidence. Points subtracted are in parentheses. Publication bias was rated as “not likely,” “likely,” or “very likely” [23].

^aWhat do these results mean given 10% disease prevalence among individuals being screened for TB?

^bOutcomes were ranked by their relative importance as critical, important, or of limited importance. Ranking helped to focus attention on those outcomes that were considered most important.

^cOnly 14/28 (50%) studies were considered to include a representative patient population; 75% of studies reported blinding of the serological test result.

^dWe downgraded for indirectness because these studies were used as a proxy for a diagnostic strategy using serological tests in addition to smear microscopy.

^eThere was considerable heterogeneity in study results.

^fWe did not pool accuracy estimates. The 95% CIs were wide for many individual studies. We did not downgrade as there were a large number of studies and we had already taken off two points for inconsistency.

^gData included in the review did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests. Therefore, publication bias cannot be ruled out. It is prudent to assume some degree of publication bias as studies showing poor performance of serological tests were probably less likely to be published. No points were deducted.

doi:10.1371/journal.pmed.1001062.t004

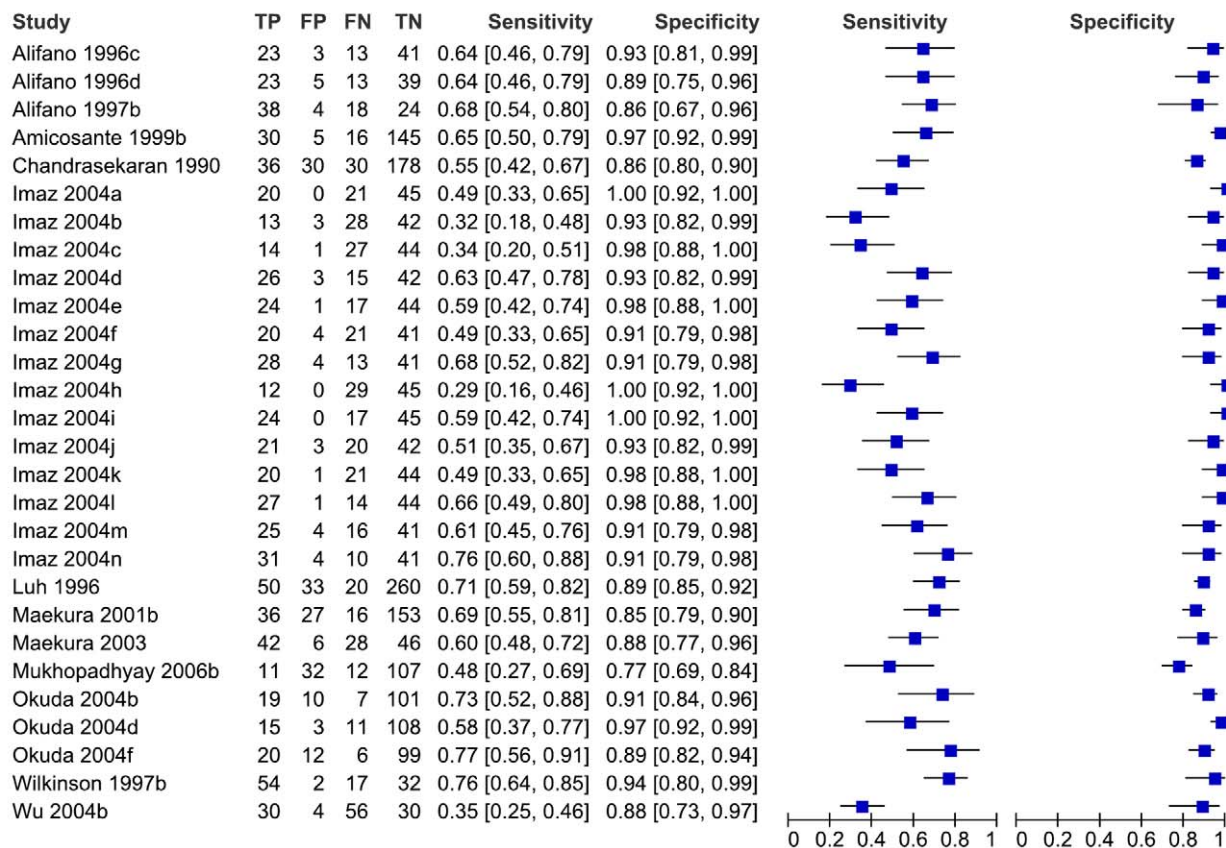


Figure 5. Forest plots of sensitivity and specificity, pulmonary TB, studies of smear-negative patients. FN, false negative; FP, false positive; TN, true negative; TP, true positive. 95% CIs are included in the brackets. On the right, the sensitivity and specificity estimates for individual studies are indicated by blue squares, and the 95% CIs are indicated by black horizontal lines.
doi:10.1371/journal.pmed.1001062.g005

test accuracy across studies. Findings from the current review are similar to those of the previous review: studies of current serological tests show that these tests provide inaccurate and imprecise estimates of sensitivity and specificity.

In the earlier systematic reviews, we recommended the use of guidelines such as STARD (Standards for the Reporting of Diagnostic Accuracy Studies) [70] and QUADAS [21] to improve methodological study quality. In the current review, within-study quality continues to be a concern. For example, in the pulmonary TB group there were 16 new studies. Six of these studies (three papers) were published subsequent to the previous reviews [20,37,49]. Four of the six studies selected participants by convenience or did not report the manner of selection (selection bias), and no studies reported that the serological test result was interpreted without knowledge of the reference standard. Selection bias and absence of blinding are features of study design that have been associated with exaggerated accuracy estimates [71,72].

A substantial contribution of the current review is the use of the GRADE approach. This framework enabled us to synthesize data on the quality of the body of evidence in a way that was not possible for the previous systematic reviews [12,13] because GRADE was not well developed for diagnostic studies at that time. The very low quality of evidence for the studies evaluating anti-TB IgG in smear-negative patients decreases our confidence in the pooled sensitivity and specificity estimates. In this subgroup, applying the GRADE approach, quality was compromised by three factors: (1) risk of bias: no studies recruited participants in a random or consecutive

manner, and only one study reported blinded interpretation of the serological test result; (2) indirectness: no studies were conducted in low- or middle-income countries, limiting generalizability to these settings; and (3) imprecision. If the pooled estimates of test accuracy had been derived from high-quality studies, then the serological test might have been shown to have some clinical utility for contributing to diagnostic algorithms for smear-negative TB, especially since the tests are relatively inexpensive, rapid, and easy to perform. However, the very low quality of the evidence implies that the serological test cannot be recommended.

Strengths and Limitations

Strengths of our review include the use of a standard protocol and comprehensive search strategy, two independent reviewers at all stages of the review process, the assessment of methodological quality of individual studies with the QUADAS tool, and the use of the GRADE approach. Heterogeneity is to be expected in results of diagnostic test accuracy studies [73]. Therefore, we prespecified subgroups to limit heterogeneity and, as noted above, used a bivariate random effects model. Our review also had limitations, notably, the majority of studies were not considered to include patients with a representative spectrum of disease severity. Differing criteria for patient selection and greater duration and severity of illness of the study populations may have introduced variability in findings among studies. In addition, the majority of studies were not performed in a blinded manner, or blinding was not explicitly stated. Also, the meta-analysis was limited by the small number of

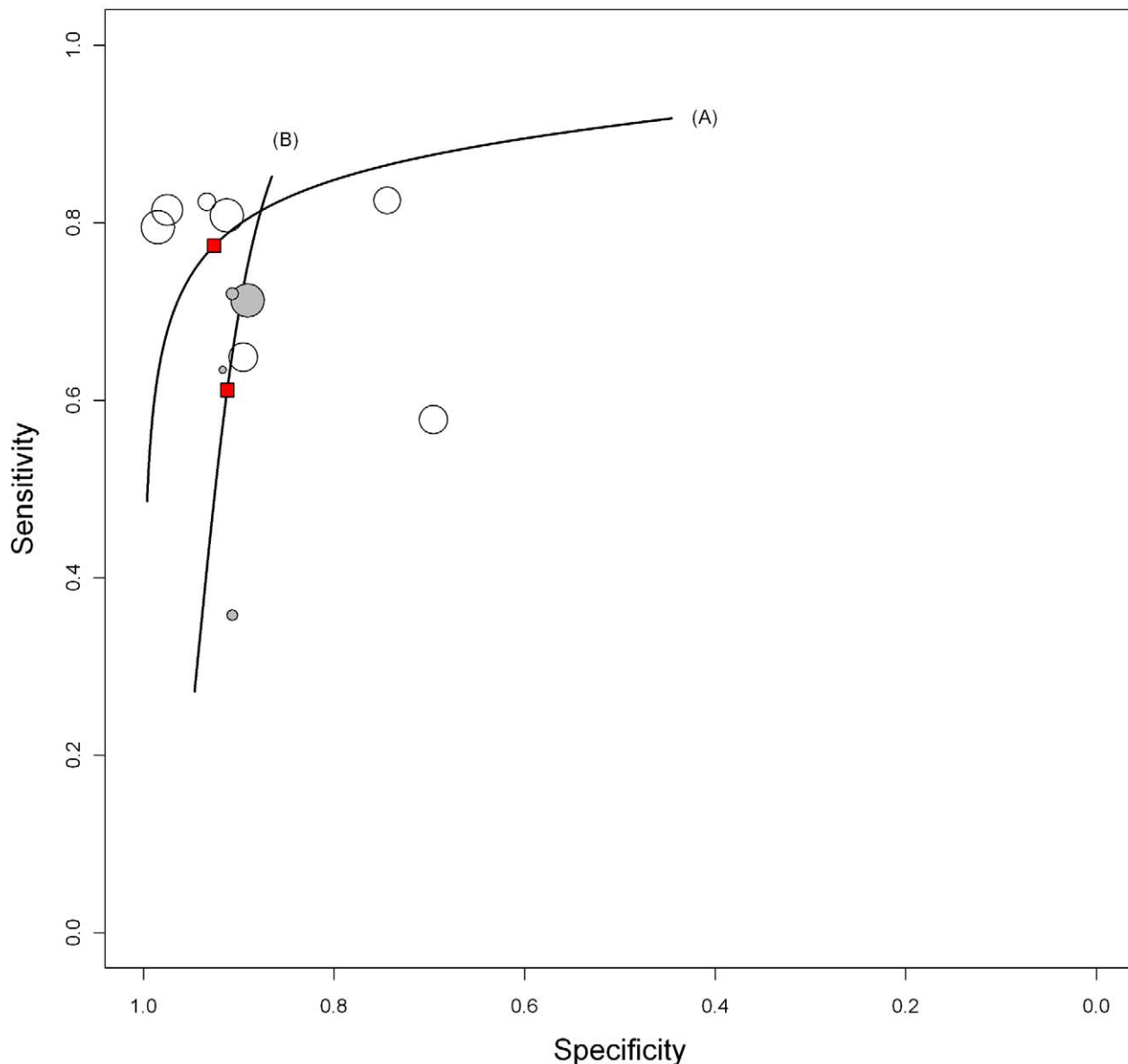


Figure 6. Summary HSROC plots of sensitivity and specificity for anda-TB IgG in smear-positive and smear-negative pulmonary TB patients. Smear-positive (line A; open circles) and smear-negative (line B; gray circles) pulmonary TB patients. The width of the circles is proportional to the number of patients in each study. The red squares are the summary values for sensitivity and specificity.
doi:10.1371/journal.pmed.1001062.g006

studies for a particular serological test. anda-TB IgG was the only test with enough studies for meta-analysis. Clearly, having more studies would have allowed us to examine observed, study-level covariates that could be sources of heterogeneity. An additional limitation was that, in some cases, we assumed that multiple results carried out on the same sample were independent. By doing so, our meta-analysis model may have underestimated heterogeneity and overestimated precision of the pooled sensitivity and specificity estimates by including a larger number of participants. Subgroup analyses in a meta-analysis, like subgroup analyses in a clinical trial, are vulnerable to bias; therefore, the findings of this meta-analysis should be interpreted with caution [74]. Although we tried to address language bias by performing the updated literature search

in all languages, the original literature search was limited to studies published in English, and language bias remains a possibility. Finally, our review did not allow for formal assessment of publication bias using methods such as funnel plots or regression tests because such techniques have not been adequately evaluated for diagnostic data [75]. Therefore, publication bias cannot be ruled out. However, it is prudent to assume some degree of publication bias, as studies showing poor performance of serological tests may have been less likely to be published, especially because several studies were industry supported.

This systematic review focused on test accuracy (i.e., sensitivity and specificity). Although, we looked for information on patient-important outcomes (meaning a serological test used

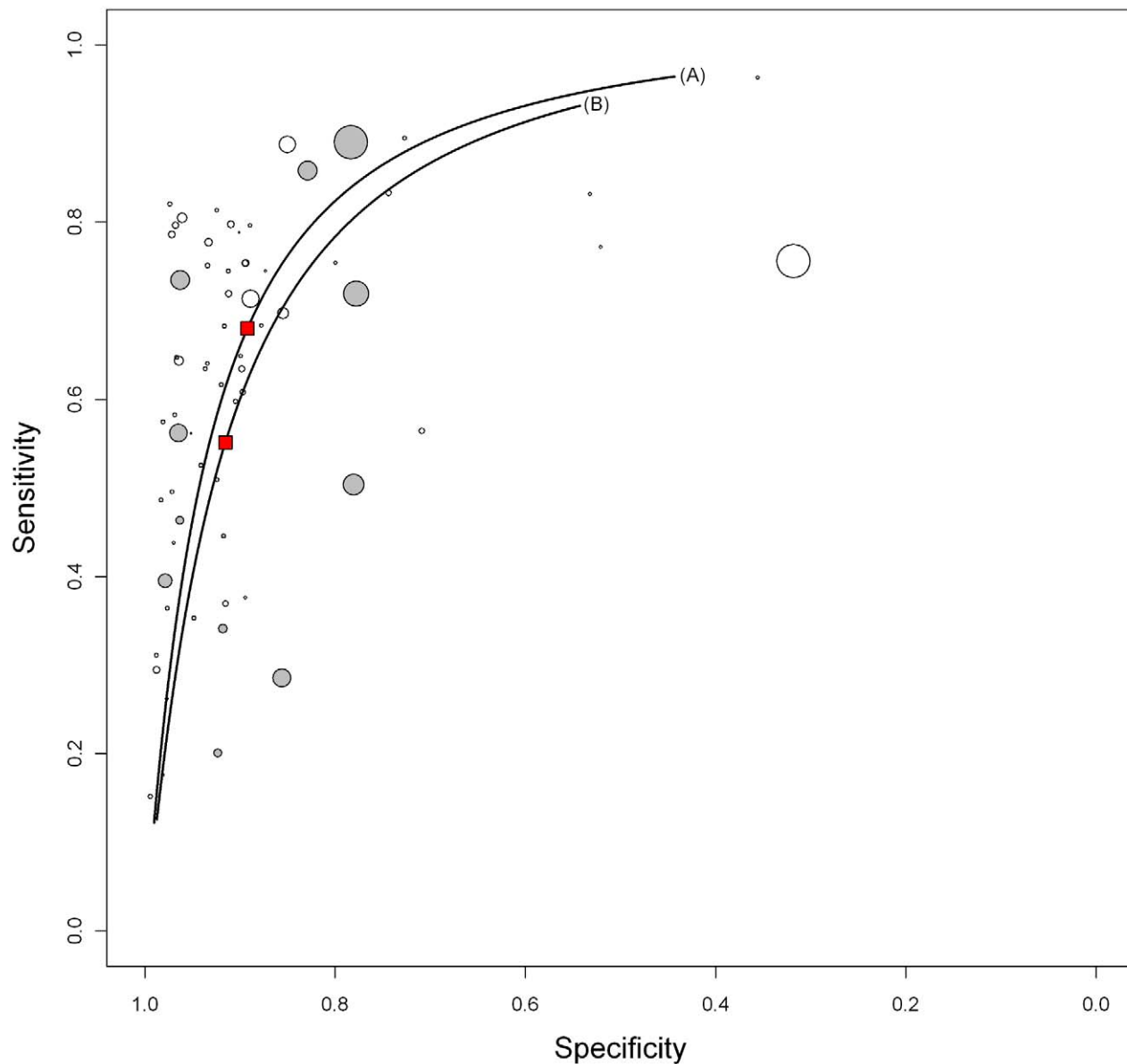


Figure 7. Summary HSROC plots of sensitivity and specificity by assay technique. ELISA (line A; open circles) and immunochromatographic test (line B; gray circles). The width of the circles is proportional to the number of patients in each study. The red squares are the summary values for sensitivity and specificity.

doi:10.1371/journal.pmed.1001062.g007

in a given situation results in a clinically relevant improvement in patient care and/or outcomes), we did not find this information in the literature reviewed. We did not identify studies with the specific aim of detecting the value of serology over and above conventional tests such as smears. However, the WHO Special Programme for Research and Training in Tropical Diseases report on rapid serological tests for TB mentioned above did evaluate the added value of smear plus serology and reported a gain equivalent to the detection of 57% of the smear-negative, culture-positive TB cases. There was, however, a corresponding unacceptable decrease in specificity to 58% [15].

In conclusion, published data on commercial serological tests produce inconsistent and imprecise estimates of sensitivity and specificity, and the quality of the body of evidence on these

tests remains disappointing. This systematic review included evaluations of only commercially available antibody-based detection tests. Considerable research is underway on new approaches to the serological diagnosis of TB. These approaches include the use of newly identified selected purified recombinant antigens and antigen combinations [14]. Recent studies from a number of laboratories have reported several new potential candidate antigens that may be expected to lead to improved antibody detection tests for TB in the future. These conclusions should be reconsidered if, in the future, methodologically adequate research evaluating serological tests becomes available.

The findings from this systematic review were used as the input for a cost-effectiveness study of serological testing for active TB in India [76]. In comparison with sputum

	Representative spectrum?	Acceptable reference standard?	Acceptable delay between tests?	Partial verification avoided?	Differential verification avoided?	Incorporation avoided?	Reference standard results blinded?	Index test results blinded?	Relevant clinical information?	Uninterpretable results reported?	Withdrawals explained?	Conducted without conflict of interest?
Alifano 1998a	+	+	+	+	+	+	?	?	+	?	?	?
Alifano 1998b	+	+	+	+	?	+	?	?	+	?	?	?
Banerjee 2003a	+	+	+	+	+	+	?	?	+	?	?	+
Banerjee 2003b	+	+	+	+	+	+	?	?	+	?	?	+
Caminero 1993	?	+	+	+	+	+	?	?	+	?	?	?
Caminero 1994	?	+	+	+	?	+	?	?	+	?	?	?
Chierakul 2001	?	+	+	+	+	+	?	+	+	?	+	+
Gevaudan 1992a	+	+	+	+	?	+	+	?	+	?	?	?
Gevaudan 1992b	+	+	+	+	?	+	+	?	+	?	?	?
Gevaudan 1992c	+	+	+	+	?	+	+	?	+	?	?	?
Gevaudan 1992d	+	+	+	+	?	+	+	?	+	?	?	?
Gevaudan 1992e	+	+	+	+	?	+	+	?	+	?	?	?
Gevaudan 1992f	+	+	+	+	?	+	+	?	+	?	?	?
Gevaudan 1992g	+	+	+	+	?	+	+	?	+	?	?	?
Gevaudan 1992h	+	+	+	+	?	+	+	?	+	?	?	?
Kunter 2003a	?	+	+	+	?	+	+	?	+	?	?	?
Kunter 2003b	?	+	+	+	?	+	+	?	+	?	?	?
Kunter 2003c	?	+	+	+	?	+	+	?	+	?	?	?
Luh 1996	?	+	+	+	?	+	+	?	+	?	?	?
McConkey 2002	?	+	+	+	+	+	+	+	+	+	+	+
Nsanze 1997a	+	+	+	+	?	+	?	?	+	?	?	?
Nsanze 1997b	+	+	+	+	?	+	?	?	+	?	?	?
Ratanasuwan 1997a	?	+	+	+	+	+	?	+	+	?	+	?
Ratanasuwan 1997b	?	+	+	+	+	+	?	+	+	?	+	?
Senol 2007	+	+	+	+	+	+	?	?	+	?	?	?

Figure 8. Methodological quality summary, all studies, extrapulmonary TB. Review authors' judgments about each methodological quality item. doi:10.1371/journal.pmed.1001062.g008

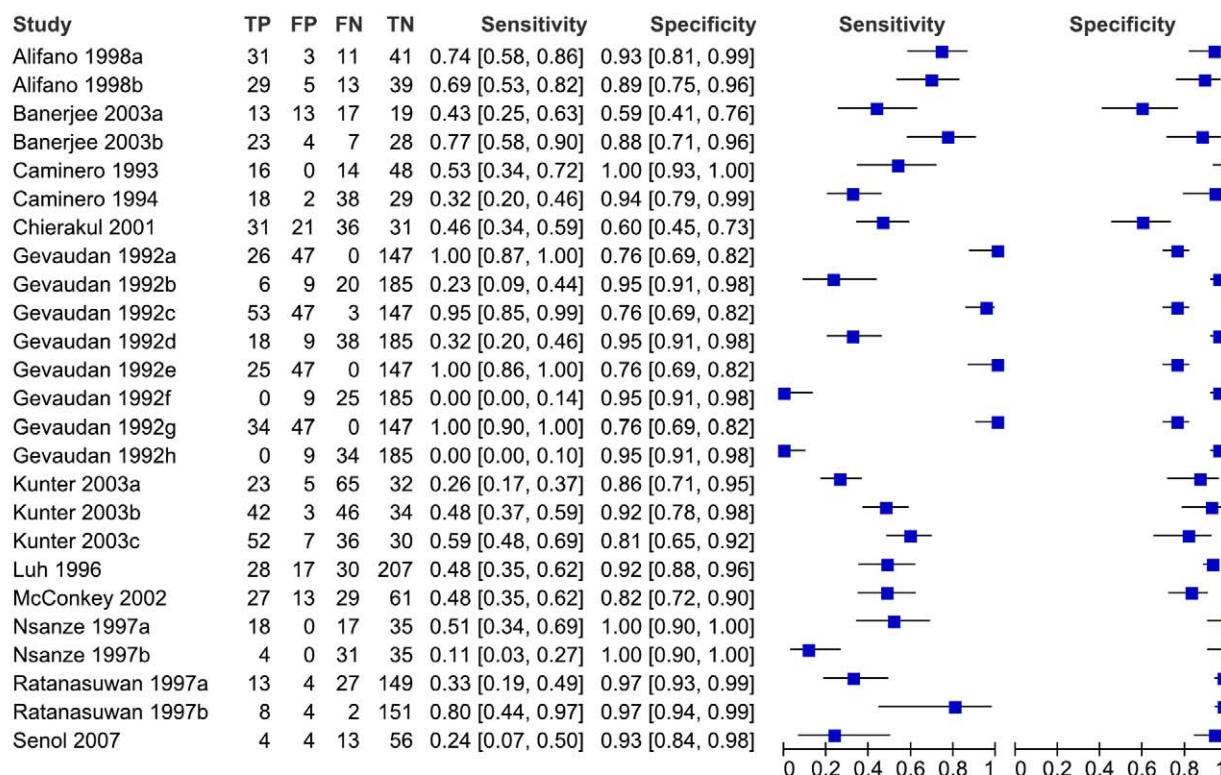


Figure 9. Forest plots of sensitivity and specificity, all studies, extrapulmonary TB. FN, false negative; FP, false positive; TN, true negative; TP, true positive. 95% CIs are included in the brackets. On the right, the sensitivity and specificity estimates for individual studies are indicated by blue squares, and the 95% CIs are indicated by black horizontal lines. doi:10.1371/journal.pmed.1001062.g009

microscopy, serological testing resulted in fewer disability-adjusted life years averted and more false-positive diagnoses and secondary infections, while increasing costs to the Indian TB control sector approximately 4-fold. This cost-effectiveness study and the findings from our updated systematic review were considered by a WHO Expert Group on Serodiagnostics, and in July 2011, the WHO published a policy statement on commercial serodiagnostic tests for diagnosis of TB. The policy states that “Commercial serological tests provide inconsistent and imprecise estimates of sensitivity and specificity. There is no evidence that existing commercial serological assays improve patient-important outcomes, and high proportions of false-positive and false-negative results adversely impact patient safety. Overall data quality was graded as very low, with harms/risks far outweighing any potential benefits (strong recommendation). It is therefore recommended that these tests should not be used in individuals suspected of active pulmonary or extra-pulmonary TB, irrespective of their HIV status.” The WHO policy strongly encourages targeted further research to identify new/alternative point-of-care tests for TB diagnosis and/or serological tests with improved accuracy [77].

Supporting Information

Figure S1 Methodological quality summary, studies of anda-TB IgG, smear-positive patients. Review authors’ judgments about each methodological quality item. (TIF)

Figure S2 Methodological quality summary, studies of anda-TB IgG, smear-negative patients. Review authors’ judgments about each methodological quality item. (TIF)

Table S1 TB serological assays on the Indian market with accuracy estimates from package inserts. (DOC)

Table S2 Characteristics of included studies evaluating serological tests for the diagnosis of pulmonary TB. (DOC)

Table S3 Characteristics of included studies evaluating serological tests for the diagnosis of extrapulmonary TB. (DOC)

Text S1 PRISMA statement. (DOC)

Text S2 Protocol. (DOC)

Text S3 PubMed literature search, 1 May 2006 to 10 February 2010. A weekly search was run through 29 June 2010. (DOC)

Text S4 List of excluded studies and reasons for their exclusion. (DOC)

Acknowledgments

We wish to thank the following individuals for their contributions to this project: Jane Cunningham (WHO Special Programme for Research and Training in Tropical Diseases), Chris Gilpin (WHO Stop TB Department), Krystal Kobasic (University of California, San Francisco), Megan Henry (Sacramento, California), Christian Lienhardt (Stop TB Partnership Research Movement), Anna Meddaugh (Portland, Oregon), Dick Menzies (McGill University), Alan Mishchenko (University of California, Berkeley), Carl-Michael Nathanson (WHO Special Programme for Research and Training in Tropical Diseases), Marek Perkowski (Portland State University), John Phillips (Fishbone Memorial Library, University of California, San Francisco), Irina Rudoy (University of California, San Francisco), Holger Schünemann (McMaster University), Karin Weldingh (Novo Nordisk), Karin Weyer (WHO Stop TB Department), Gloria Won (Fishbone Memorial Library, University of California, San Francisco), and

George Yen (University of California, San Francisco). We thank David Dowdy (University of California, San Francisco) and the reviewers for their helpful comments on the manuscript. In addition, we are grateful to Mien Pathy and Jia Si Duan for administrative support and the members of the WHO Expert Group Meeting on Serodiagnostics, July 2010.

Author Contributions

Conceived and designed the experiments: KRS AR MP. Performed the experiments: KRS LFF. Analyzed the data: KRS ND IS. Contributed reagents/materials/analysis tools: ND IS. Wrote the paper: KRS MP. ICMJE criteria for authorship read and met: KRS LFF ND IS SL AR PCH MP. Agree with the manuscript's results and conclusions: KRS LFF ND IS SL AR PCH MP. Wrote the first draft of the paper: KRS. Obtained funding: AR.

References

- Lonnroth K, Castro KG, Chakaya JM, Chauhan LS, Floyd K, et al. (2010) Tuberculosis control and elimination 2010–50: cure, care, and social development. *Lancet* 375: 1814–1829.
- Small PM, Pai M (2010) Tuberculosis diagnosis—time for a game change. *N Engl J Med* 363: 1070–1071.
- Harries A (2004) How does the diagnosis of tuberculosis in persons infected with HIV differ from diagnosis in persons not infected with HIV? In: Frieden T, ed. *Toman's tuberculosis: case detection, treatment, and monitoring—questions and answers*, 2nd ed. Geneva: World Health Organization, pp 80–83.
- Perkins MD, Cunningham J (2007) Facing the crisis: improving the diagnosis of tuberculosis in the HIV era. *J Infect Dis* 196(Suppl 1): S15–S27.
- 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.
- Boehme CC, Nabeta P, Hillebrand M, Nicol MP, Shenai S, et al. (2010) Rapid molecular detection of tuberculosis and rifampin resistance. *N Engl J Med* 363: 1005–1015.
- World Health Organization (2010) Roadmap for rolling out Xpert MTB/RIF for rapid diagnosis of TB and MDR-TB Geneva: World Health Organization. Available: http://www.who.int/tb/laboratory/roadmap_xpert_mtb-rif.pdf. Accessed 6 March 2011.
- Hopewell PC, Pai M, Maher D, Uplekar M, Raviglione MC (2006) International standards for tuberculosis care. *Lancet Infect Dis* 6: 710–725.
- World Health Organization Special Programme for Research and Training in Tropical Diseases, Foundation for Innovative New Diagnostics (2006) Diagnostics for tuberculosis: global demand and market potential. Geneva: World Health Organization, Available: <http://apps.who.int/tdr/svc/publications/tdr-research-publications/diagnostics-tuberculosis-global-demand>. Accessed 20 June 2010.
- Grenier J, Pinto L, Nair D, Steingart KR, Dowdy D, et al. (2011) Widespread use of serological tests for tuberculosis: data from 22 high-burden countries. *Eur Respir J*. In press.
- Dinnes J, Deeks J, Kunst H, Gibson A, Cummins E, et al. (2007) A systematic review of rapid diagnostic tests for the detection of tuberculosis infection. *Health Technol Assess* 11: 1–196.
- Steingart KR, Henry M, Laal S, Hopewell PC, Ramsay A, et al. (2007) Commercial serological antibody detection tests for the diagnosis of pulmonary tuberculosis: a systematic review. *PLoS Med* 4: e202. doi:10.1371/journal.pmed.0040202.
- 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* 62: 911–918.
- Steingart KR, Dendukuri N, Henry M, Schiller I, Nahid P, et al. (2009) Performance of purified antigens for serodiagnosis of pulmonary tuberculosis: a meta-analysis. *Clin Vaccine Immunol* 16: 260–276.
- Special Programme for Research and Training in Tropical Diseases (2008) Laboratory-based evaluation of 19 commercially available rapid diagnostic tests for tuberculosis. Geneva: World Health Organization.
- Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM (2008) Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 149: 889–897.
- Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 6: e1000097. doi:10.1371/journal.pmed.1000097.
- Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi YT (2010) Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. *Cochrane handbook for systematic reviews of diagnostic test accuracy*, version 0.9.0. London: The Cochrane Collaboration, Available: <http://srdta.cochrane.org/> Accessed 6 March 2011.
- World Bank (2010) World Bank list of economies. Washington (District of Columbia): World Bank, Available: <http://siteresources.worldbank.org/DATASTATISTICS/Resources/CLASS.XLS>. Accessed 5 August 2010.
- Anderson BL, Welch RJ, Litwin CM (2008) Assessment of three commercially available serologic assays for detection of antibodies to *Mycobacterium tuberculosis* and identification of active tuberculosis. *Clin Vaccine Immunol* 15: 1644–1649.
- 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.
- Pai M, Minion J, Steingart K, Ramsay A (2010) New and improved tuberculosis diagnostics: evidence, policy, practice, and impact. *Curr Opin Pulm Med* 16: 271–284.
- Schunemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, et al. (2008) Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 336: 1106–1110.
- Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, et al. (2010) GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 64: 401–406.
- Rutter CM, Gatsonis CA (2001) A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 20: 2865–2884.
- Spiegelhalter D, Thomas A, Best N, Lunn D (2004) WinBUGS user manual, version 1.4.1. Available: <http://www.mrc-bsu.cam.ac.uk/bugs>. BUGS Project. Accessed 13 May 2010.
- Alan A, Coull A (1998) Approximate is better than “exact” for interval estimation of binomial proportions. *Am Stat* 52: 119–126.
- Jones CM, Athanasiou T (2005) Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. *Ann Thorac Surg* 79: 16–20.
- R Development Core Team (2006) R: a language and environment for statistical computing. Available: <http://www.R-project.org>. R Project for Statistical Computing. Accessed 13 May 2010.
- 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.
- 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.
- 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.
- 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.
- Bukhary ZA (2007) Evaluation of anti-A60 antigen IgG enzyme-linked immunosorbent assay for serodiagnosis of pulmonary tuberculosis. *Ann Thorac Med* 2: 47–51.
- 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.
- 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.
- D'Alessandro A, de Waard JH (2008) [Evaluation of two commercial tests for the serodiagnosis of pulmonary tuberculosis.] *Rev Chilena Infectol* 25: 37–40.
- Erer OF, Biçmen C, Kıraklı C, Aktođu S, Florat N, et al. (2001) [Diagnostic value of a rapid immunochromatographic test (ICT) in patients with active pulmonary tuberculosis.] *Toraks Dergisi* 2: 80–84.
- Gao MQ, Chu NH, Wang HY, Zhao QR, Li H, et al. (2007) [The clinical significance of serum tuberculosis specific antigen antibody in the diagnosis of tuberculosis.] *Zhonghua Jie He He Hu Xi Za Zhi* 30: 918–920.
- 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.

41. 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.
42. 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.
43. 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.
44. Kassa-Kelembho E, Kassa E, Zandanga G, Service YB, Ignaleamoko A, et al. (2006) Poor performance of a novel serological test for diagnosis of pulmonary tuberculosis in Bangui, Central African Republic. *Clin Vaccine Immunol* 13: 702–703.
45. 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.
46. Mackura R, Kohno H, Hirotsu 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.
47. Mackura 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.
48. 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.
49. Mizusawa M, Kawamura M, Takamori M, Kashiyama T, Fujita A, et al. (2008) Increased synthesis of anti-tuberculous glycolipid immunoglobulin G (IgG) and IgA with cavity formation in patients with pulmonary tuberculosis. *Clin Vaccine Immunol* 15: 544–548.
50. Mukhopadhyay A, Guan M, Chen HY, Lu Y, Lim TK (2006) Prospective study of a new serological test (ASSURE TB Rapid Test) for the diagnosis of pulmonary tuberculosis. *Int J Tuberc Lung Dis* 10: 620–624.
51. Nurkic M, Imamovic A, Numanovic F, Gegic M, Osmic M, et al. (2006) [Diagnostical values of antibody on A60 antigen in diagnostic of tuberculosis.] *Med Arh* 60: 166–170.
52. Okuda Y, Mackura R, Hirotsu 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.
53. 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.
54. Platonova IL, Sakhelashvili MI (2007) [Comparative evaluation of the informative value of the rapid serochek mbt test to determine mycobacterium tuberculosis antibodies and mantoux tuberculin test in adults.] *Probl Tuberk Bolezn Legk* 2007: 10–13.
55. 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. *Inter J Tuberc Lung Dis* 3: 231–238.
56. Traunmüller 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.
57. 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.
58. Wu HP, Hua CC, Yu CC, Wu SY (2005) Comparison of plasma interferon-gamma and antigen 60 immunoglobulin G in diagnosing pulmonary Mycobacterium tuberculosis infection. *Chang Gung Med J* 28: 779–785.
59. 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.
60. Alifano M, De Pascalis R, Sofia M, Faraone S, Del Pezzo M, et al. (1998) Detection of IgG and IgA against the mycobacterial antigen A60 in patients with extrapulmonary tuberculosis. *Thorax* 53: 377–380.
61. Banerjee S, Gupta S, Shende N, Kumar S, Harinath BC (2003) Serodiagnosis of tuberculosis using two ELISA systems. *Indian J Clin Biochem* 18: 48–53.
62. Caminero JA, Rodriguez de Castro F, Carrillo T, Diaz F, Rodriguez Bermejo JC, et al. (1993) Diagnosis of pleural tuberculosis by detection of specific IgG anti-antigen 60 in serum and pleural fluid. *Respiration* 60: 58–62.
63. Caminero JA, Rodriguez de Castro F, Carrillo T, Lafarga B, Diaz F, et al. (1994) Value of ELISA using A60 antigen in the serodiagnosis of tuberculosis. *Respiration* 61: 283–286.
64. Chierakul N, Damrongchokpipat P, Chairprasert A, Arjratanakul W (2001) Antibody detection for the diagnosis of tuberculous pleuritis. *Int J Tuberc Lung Dis* 5: 968–972.
65. Gevaudan MJ, Bollet C, Charpin D, Mallet MN, De Micco P (1992) Serological response of tuberculosis patients to antigen 60 of BCG. *Eur J Epidemiol* 8: 666–676.
66. 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.
67. Nsanze H, Ameen AS, Fares E, Vargees L, Mustafa N (1997) Serodiagnosis of tuberculosis and leprosy by enzyme immunoassay. *Clin Microbiol Infect* 3: 236–239.
68. Ratanasuwan W, Kreiss JK, Nolan CM, Schaeffler BA, Suwanagool S, et al. (1997) Evaluation of the MycoDot test for the diagnosis of tuberculosis in HIV seropositive and seronegative patients. *Int J Tuberc Lung Dis* 1: 259–264.
69. Senol G, Erer OF, Yalcin YA, Coskun M, Gunduz AT, et al. (2007) Humoral immune response against 38-kDa and 16-kDa mycobacterial antigens in tuberculosis. *Eur Respir J* 29: 143–148.
70. 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.
71. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, et al. (1999) Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 282: 1061–1066.
72. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, et al. (2006) Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 174: 469–476.
73. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA (2007) A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 8: 239–251.
74. Davey Smith G, Egger M, Phillips AN (1997) Meta-analysis. Beyond the grand mean? *BMJ* 315: 1610–1614.
75. 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.
76. Dowdy DW, Steingart KR, Pai M (2011) Serological testing versus other strategies for diagnosis of active tuberculosis in India: A cost-effectiveness analysis. *PLoS Med* 8: e1001074. doi:10.1371/journal.pmed.1001074.
77. World Health Organization (2011) Policy Statement: Commercial serodiagnostic tests for diagnosis of tuberculosis. WHO, Geneva, Switzerland. WHO/HTM/TB/2011.5. Available: http://www.who.int/tb/laboratory/policy_statement/en. Accessed 12 July 2011.

Editors' Summary

Background Every year nearly 10 million people develop tuberculosis—a contagious bacterial infection—and about two million people die from the disease. *Mycobacterium tuberculosis*, the bacterium that causes tuberculosis, is spread in airborne droplets when people with the disease cough or sneeze. It usually infects the lungs (pulmonary tuberculosis) but can also infect the lymph nodes, bones, and other tissues (extrapulmonary tuberculosis). The characteristic symptoms of tuberculosis are a persistent cough, weight loss, and night sweats. Diagnostic tests for the disease include microscopic examination of sputum (mucus brought up from the lungs by coughing) for *M. tuberculosis* bacilli, chest radiography, mycobacterial culture (in which bacteriologists try to grow *M. tuberculosis* from sputum or tissue samples), and nucleic acid amplification tests (which detect the bacterium's genome in patient samples). Tuberculosis can usually be cured by taking several powerful drugs daily or several times a week for at least six months.

Why Was This Study Done? Although efforts to control tuberculosis have advanced over the past decade, missed tuberculosis diagnoses and mismanaged tuberculosis continue to fuel the global epidemic. A missed diagnosis may lead to more severe illness and death, especially for people infected with both tuberculosis and HIV. Also, a missed diagnosis means that an untreated individual with pulmonary tuberculosis may remain infectious for longer, continuing to spread tuberculosis within the community. Missed diagnoses are a particular problem in resource-limited countries where sputum microscopy and chest radiography often perform poorly and other diagnostic tests are too expensive and complex for routine use. Serological tests, which detect antibodies against *M. tuberculosis* in the blood (antibodies are proteins made by the immune system in response to infections), might provide a way to diagnose tuberculosis in resource-limited countries. Indeed, many serological tests for tuberculosis diagnosis are on sale in developing countries. However, because of doubts about the accuracy of these commercial tests, they are not recommended for use in routine practice. In this systematic review and meta-analysis, the researchers assess the diagnostic accuracy of commercial serological tests for pulmonary and extrapulmonary tuberculosis. A systematic review uses predefined criteria to identify all the research on a given topic; meta-analysis is a statistical method that combines the results of several studies.

What Did the Researchers Do and Find? The researchers searched the literature for studies that evaluated serological tests for active tuberculosis published between 1990 and 2010. They used data from these studies to calculate each test's sensitivity (the proportion of patients with a positive serological test among patients with tuberculosis confirmed by a reference method; a high sensitivity indicates that the test detects most patients with tuberculosis) and specificity (the proportion of patients with a negative serological result among people without tuberculosis; a high specificity means the test gives few false-positive diagnoses). They also assessed the methodological quality of each study and rated the overall quality of the evidence. The researchers found 67 studies (half from low/middle-income countries)

that evaluated serological tests for the diagnosis of pulmonary tuberculosis. The sensitivity of these tests varied between studies, ranging from 0% to 100%; their specificities ranged from 31% to 100%. For the anda-TB IgG test—the only test with sufficient studies for a meta-analysis—the pooled sensitivity from the relevant studies was 76% in smear-positive patients and 59% in smear-negative patients. The pooled specificities were 92% and 91%, respectively. The researchers found 25 studies (40% from low/middle-income countries) that evaluated serological tests for the diagnosis of extrapulmonary tuberculosis. Again, sensitivities and specificities for each test varied greatly between studies, ranging from 0% to 100% and 59% to 100%, respectively. Overall, for both pulmonary and extrapulmonary tuberculosis, the quality of evidence from the studies of the serological tests was graded very low.

What Do These Findings Mean? This systematic review, which updates an analysis published in 2007, indicates that commercial serological tests do not provide an accurate diagnosis of tuberculosis. This finding confirms previous systematic reviews of the evidence, despite a recent expansion in the relevant literature. Moreover, the researchers' analysis indicates that the overall quality of the body of evidence on these tests remains poor. Many of the identified studies used unsatisfactory patient selection methods, for example. Clearly, there is a need for continued and improved research on existing serological tests and for research into new approaches to the serological diagnosis of tuberculosis. For now, though, based on these findings, cost-effectiveness data, and expert opinion, the World Health Organization has issued a recommendation against the use of currently available serological tests for the diagnosis of tuberculosis, while stressing the importance of continued research on these and other tests that could provide quick and accurate diagnosis of TB.

Additional Information Please access these Web sites via the online version of this summary at <http://dx.doi.org/10.1371/journal.pmed.1001062>.

- The World Health Organization provides information on all aspects of tuberculosis, including information on tuberculosis diagnostics on the Stop TB Partnership (some information is in several languages); the Strategic and Technical Advisory Group for Tuberculosis recommendations on tuberculosis diagnosis are available
- The Web site Evidence-Based Tuberculosis Diagnosis (from Stop TB Partnership's New Diagnostics Working Group) provides access to several resources on TB diagnostics, including systematic reviews, guidelines, and training materials
- The US Centers for Disease Control and Prevention has information about tuberculosis, including information on the diagnosis of tuberculosis disease
- The US National Institute of Allergy and Infectious Diseases also has detailed information on all aspects of tuberculosis
- MedlinePlus has links to further information about tuberculosis (in English and Spanish)