Nucleic acid amplification tests in the diagnosis of tuberculous pleuritis: a systematic review and meta-analysis

Running title: Nucleic acid amplification tests for tuberculous pleuritis

Madhukar Pai\textsuperscript{1}, Laura L. Flores\textsuperscript{2}, Alan Hubbard\textsuperscript{3}, Lee W. Riley\textsuperscript{2}, John M. Colford Jr.\textsuperscript{1}

\textsuperscript{1}Divisions of Epidemiology, \textsuperscript{2}Infectious Diseases, and \textsuperscript{3}Biostatistics

School of Public Health
University of California, Berkeley
140, Warren Hall, Berkeley, CA 94720, USA

Email: Madhukar Pai – madhupai@uclink.berkeley.edu; Laura Flores - floresll@uclink.berkeley.edu; Alan Hubbard - hubbard@stat.berkeley.edu; Lee Riley - lwriley@uclink4.berkeley.edu; John M Colford - jcolford@socrates.berkeley.edu

Corresponding author:

John M. Colford, MD, PhD
Associate Professor
University of California, Berkeley
School of Public Health
Division of Epidemiology
140, Warren Hall
Berkeley CA 94720

Email: jcolford@socrates.berkeley.edu
Telephone: 510-643-1076
Fax: 413-228-5931
ABSTRACT

**Background:** Conventional tests for tuberculous pleuritis have several limitations. A variety of new, rapid tests such as nucleic acid amplification tests—including polymerase chain reaction—have been evaluated in recent times. We conducted a systematic review to determine the accuracy of nucleic acid amplification tests in the diagnosis of tuberculous pleuritis.

**Methods:** A systematic review and meta-analysis of 38 English and Spanish articles (with 40 studies), identified via searches of six electronic databases, hand searching of selected journals, and contact with authors, experts, and test manufacturers. Sensitivity, specificity, and other measures of accuracy were pooled using random effects models. Summary receiver operating characteristic curves were used to summarize overall test performance. Heterogeneity in study results were formally explored using subgroup analyses.

**Results:** Of the 40 studies included, 26 used in-house (“home-brew”) tests, and 14 used commercial tests. Commercial tests had a low overall sensitivity (0.62; 95% confidence interval [CI] 0.43, 0.77), and high specificity (0.98; 95% CI 0.96, 0.98). The positive and negative likelihood ratios for commercial tests were 25.4 (95% CI 16.2, 40.0) and 0.40 (95% CI 0.24, 0.67), respectively. All commercial tests had consistently high specificity estimates; the sensitivity estimates, however, were heterogeneous across studies. With the in-house tests, both sensitivity and specificity estimates were significantly heterogeneous. Clinically meaningful summary estimates could not be determined for in-house tests.
Conclusions: Our results suggest that commercial NAA tests may have a potential role in confirming tuberculous pleuritis. However, these tests have low and variable sensitivity and, therefore, may not be useful in excluding the disease. NAA test results, therefore, cannot replace conventional tests; they need to be interpreted in parallel with clinical findings and results of conventional tests. The accuracy of in-house nucleic acid amplification tests is poorly defined because of heterogeneity in study results.

Keywords: tuberculous pleuritis, nucleic acid amplification tests, polymerase chain reaction, accuracy, meta-analysis
BACKGROUND

In the context of the human immunodeficiency virus (HIV) epidemic, clinicians frequently encounter extrapulmonary and disseminated forms of tuberculosis (TB) [1-3]. In the USA, nearly 20% of the TB cases are extra-pulmonary [2]. In England and Wales, 38% of the TB cases are extrapulmonary [3]. Tuberculous pleuritis is a common manifestation of extrapulmonary TB [4]. TB is the most common cause of pleural effusion in many countries [4]. For example, studies from Spain [5], Malaysia [6], and Saudi Arabia [7] showed that TB accounted for 25%, 44%, and 37% of all effusions, respectively. In the USA, the annual incidence of tuberculous pleuritis has been estimated to be about 1000 cases, and approximately one in 300 patients with TB will have tuberculous pleuritis [4,8]. The incidence of tuberculous effusions may be higher in patients with HIV infection [9].

Conventional diagnostic tests include microscopy of the pleural fluid, culture of pleural fluid, sputum, and pleural tissue, and pleural biopsy [4]. These tests have limitations. Microscopy of the pleural fluid is rarely positive (<5%) [10-12]. Culture of pleural fluid has low sensitivity (24% - 58%), and results are not available for weeks [11-13]. Biopsy of pleural tissue, and culture of biopsy material are widely held to be the best methods of confirming the diagnosis [4,10,12]. This combination may lead to the diagnosis 86% of the time [10]. Although not perfect, culture and/or biopsy, therefore, are widely considered the standard of diagnosis [4,10]. However, pleural biopsy is invasive, operator-dependent, and technically difficult (particularly in children) [14].

Because of the limitations of conventional tests, newer and rapid tests such as nucleic acid amplification (NAA) tests—including polymerase chain reaction (PCR)—
have been evaluated. Because of their high sensitivity and specificity in smear-positive respiratory specimens, these tests are now used—mainly in developed countries—for the direct detection of *M. tuberculosis* complex in respiratory specimens [15,16]. NAA tests are categorized as commercial or in-house (“home-brew”) tests. Commercial tests include the Amplified *Mycobacterium tuberculosis* Direct Test® (MTD) (Gen-Probe Inc, San Diego, CA), the Amplicor® MTB tests (Roche Molecular Systems, Branchburg, NJ), and the recently discontinued LCx® test (Abbott Laboratories, Abbott Park, IL). In the USA, the Amplicor test is licensed for use in smear-positive respiratory specimens; the MTD test is approved for smear-positive as well as smear-negative respiratory specimens [16]. No commercial test is licensed for use in non-respiratory specimens. We conducted this systematic review and meta-analysis to determine the overall accuracy of NAA tests in the diagnosis of tuberculous pleuritis, and to identify factors associated with heterogeneity of results between studies.

**METHODS**

**Search strategy**

We searched the following databases: PubMed (1985–January 2003), EMBASE (1988–2002), Web of Science (1990–2002), BIOSIS (1993–2002), Cochrane Library (2002; Issue 2), and LILACS (1990–2002). The *Journal of Clinical Microbiology*, a high-yield journal for TB diagnostic studies, was also hand-searched separately. All searches were up to date as of August 2002. The PubMed search was updated in January 2003. The search terms used were: tuberculosis, *Mycobacterium tuberculosis*, nucleic acid amplification techniques, polymerase chain reaction, sensitivity and specificity, and
accuracy. Experts in the field were contacted. Bibliographies from the included studies and relevant review articles were screened. We obtained lists of citations from companies that manufacture commercial tests. Although no language restrictions were imposed initially, for the full-text review and final analysis, our resources only permitted review of English and Spanish articles. Conference abstracts were excluded because of the limited data presented in them.

**Study selection**

Our search strategy was designed to include all published studies on NAA tests for the direct detection of *M. tuberculosis* in pleural fluid specimens. For inclusion, a study had to:

1. report a comparison of a NAA test against a reference standard, and provide data necessary for the computation of both sensitivity and specificity;
2. include at least 10 pleural fluid specimens (since studies with very few specimens are vulnerable to selection bias [17])

Studies on use of NAA tests in pleural biopsy and/or cytology specimens were excluded.

Two reviewers (MP and LLF) independently judged study eligibility while screening the citations. Disagreements were resolved by consensus. A list of excluded studies and a log of reasons for exclusion are available from the authors upon request.

**Data extraction and quality assessment**

Data extraction was performed by two reviewers. One reviewer (MP) extracted the data from all English studies. Another reviewer (LLF) extracted data from all Spanish articles. The second reviewer (LLF) also independently extracted data from a subset (36%) of the English articles, in order to determine the inter-rater agreement. The
abstracted data included methodological quality, participant characteristics, test methods, and outcome data.

Quality assessment was performed using methods adapted from two guidelines on systematic reviews of diagnostic studies [17,18]. For each study, the following quality criteria were scored as fulfilled or not: 1) Independent comparison of NAA test against reference standard; 2) Cross-sectional design (versus case-control design); 3) Blinded (single or double) interpretation of test and reference standard results; 4) Consecutive or random sampling of patients; 5) Prospective data collection; 6) Inclusion of at least 10 specimens/patients with confirmed tuberculous pleuritis. If no data on the above criteria were reported in the primary studies, we requested the information from the authors. For the purposes of analysis, responses coded as “not reported” were grouped together with “not met.” A high quality study was arbitrarily defined as that which met at least 5/6 criteria; a medium quality met 3 or 4 of the 6 criteria; and a low quality study met less than 3/6 criteria. Since discrepant analysis (where discordant results between NAA test and reference test results are resolved, post-hoc, using clinical data) may be a potential source of bias, we preferentially included unresolved data.

**Statistical analysis and data synthesis**

We used standard methods recommended for meta-analyses of diagnostic test evaluations [17-19]. Analyses were performed using Meta-Test [20], and Stata version 8 (Stata Corporation, Texas). We computed the following measures of test accuracy for each study: sensitivity [true positive rate (TPR)], specificity [1-false positive rate (FPR)], positive likelihood ratio (LR+), negative likelihood ratio (LR–), and diagnostic odds ratio (DOR). These measures were pooled using the random effects model [17-19].
Each study in the meta-analysis contributed a pair of numbers: TPR and FPR. Since TPR and FPR are not independent, we summarized their joint distribution by constructing a summary receiver operating characteristic (SROC) curve [21]. Unlike a traditional ROC plot used to explore the effect of varying thresholds (cut-points) on TPR and FPR in a single study, each data point in the SROC plot represents a separate study in the meta-analysis. The SROC curve (and area under the curve) represents the overall performance of the test, and depicts the trade off between sensitivity and specificity.

Heterogeneity in meta-analyses refers to the degree of variability in results across studies. We used the Chi-square and Fisher’s exact tests to detect statistically significant heterogeneity. Stratified (subgroup) analyses were used to identify study design and test-related factors responsible for heterogeneity in test accuracy. Studies using commercial tests were analyzed separately from those using in-house tests. Studies with commercial tests were further stratified by type of test (brand). Finally, since publication bias is of concern for meta-analyses of diagnostic studies [22], we tested for the potential presence of this bias using funnel plots and the Egger test [23].

RESULTS

Study selection

Figure 1 outlines our study selection process. Thirty eight articles [24-61] were included in the meta-analysis. Four articles were in Spanish [26,37,39,44]. Two articles reported evaluations of more than one NAA test against a common reference standard [30,57]. Each such test comparison was counted as a separate study. Thus, the total number of test comparisons (hereafter referred to as studies) was 40. Of these, 14 (35%) were studies of
commercial tests, and 26 (65%) were of in-house tests. The average (median) sample size of each study in the meta-analysis was 60 pleural specimens or subjects, with a range of 15 to 375.

**Study characteristics and quality**

The mean inter-rater agreement between the two reviewers for items in the quality checklist was 0.86. Our initial data were affected by incomplete reporting in the primary studies. We contacted the authors via email, and obtained additional data for 25/40 (63%) included studies. Tables 1 and 2 present the descriptive data from studies with commercial and in-house tests, respectively. The tables present data on study quality, along with sample size, sensitivity and specificity estimates. In the commercial tests subgroup (Table 1), 36%, 50%, and 14% of the studies were of high, medium and low quality, respectively. In the in-house tests subgroup (Table 2), 39%, 42%, and 19% of the studies were of high, medium and low quality, respectively. Tables 1 and 2 show the variability in study quality and variability in NAA protocols employed. Although a variety of reference tests (or combinations of reference tests) were used, culture alone, or a combination of culture plus biopsy/microscopy/clinical data were used in 39 of the 40 studies.

**Diagnostic accuracy of commercial tests**

Figure 2(A) displays the sensitivity and specificity estimates from each of the 14 studies using commercial tests, stratified by type (brand) of test. Almost all studies had specificity estimates close to 1.0. In contrast, sensitivity estimates were lower and heterogeneous (range 0.20-1.0). Figure 3(A) shows the SROC curve for the commercial tests. The regression line does not trace out a typical ROC curve—the curve shows no
trade-off between sensitivity and specificity. Table 3 presents the results of the meta-analysis. The summary measures of specificity and LR+ were very high and homogeneous. All other measures were highly heterogeneous.

Commercial tests use different amplification methods and target nucleic acid sequences: the Amplicor test employs PCR technology to amplify the 16s rRNA target; the LCx test employs ligase chain reaction to amplify the 38 kDa target; and the MTD test utilizes transcription-mediated amplification to amplify rRNA. To account for these differences, we further stratified commercial tests by type (brand) of test. Four of the studies evaluated the Amplicor test [28,42,54,55], four the LCx test [37,44,46,52], and six evaluated the MTD test [26,33-35,49,61]. Stratification reduced the overall heterogeneity to some extent (Table 3 and Figure 2A). Although specificity did not vary by type of kit, sensitivity did—the Amplicor test had a lower sensitivity (0.37) than the LCx (0.72) and MTD (0.77) tests. Since 5/6 studies on the MTD test evaluated the first generation MTD-1 kit, there were insufficient numbers of MTD-2 studies to determine if the enhanced MTD-2 test (licensed for use in smear-negative respiratory specimens) performed better than the MTD-1 test.

**Diagnostic accuracy of in-house tests**

Figure 2(B) displays the sensitivity and specificity estimates from each of the 26 studies using in-house tests. Both sensitivity (range 0.20-1.0) and specificity (range 0.53-1.0) estimates were highly variable. All summary measures were grossly heterogeneous (Table 3) and therefore would not be appropriately summarized. The SROC curve [Figure 3(B)] displays a ROC-type trade-off between sensitivity and specificity.
We performed stratified analyses to identify sources of heterogeneity among in-house tests. Table 4 presents two factors that appeared most strongly associated with the observed heterogeneity. Studies that employed a case-control design produced diagnostic odds ratio estimates nearly 2.4 times higher than studies that employed a cross-sectional design. Studies with PCR tests that used the IS6110 target sequence produced diagnostic odds ratio estimates 2.5 times higher than studies that used PCR tests with other targets. However, even after stratifying on study design and target sequence, considerable unexplained heterogeneity persisted in all the summary measures. The shape of the SROC curve suggested that variability in diagnostic thresholds (cut-points) across studies could partly explain the heterogeneity.

Publication bias

In the subgroup with commercial tests, the Egger test was not statistically significant (p = 0.55). However, in the in-house tests subgroup the Egger test was significant (p = 0.002), with an asymmetric funnel plot (figure 4)—evidence in favour of potential publication bias.

DISCUSSION

Since conventional tests are not always helpful in establishing a diagnosis of tuberculous pleuritis, several rapid tests and biomarkers have been evaluated: Adenosine Deaminase (ADA) [12,14,45,51,59,62], Interferon-γ (IFN-γ) [59,60,62,63], lysozyme [62], soluble interleukin 2 receptors [63], and NAA tests [24-61]. There has been an explosion of studies evaluating these rapid tests, and systematic reviews and meta-analyses are necessary to synthesize this growing body of literature. A recent meta-
analysis summarized the evidence on ADA and IFN-γ for the diagnosis of tuberculous pleuritis [64]. Both ADA and IFN-γ tests were found to be reasonably accurate at detecting tuberculous pleuritis. Our meta-analysis summarizes the evidence on accuracy of NAA tests in the diagnosis of tuberculous pleuritis.

Principal findings

The role of NAA tests has been reasonably well defined in pulmonary tuberculosis [15,16,65], and guidelines exist for testing of respiratory specimens [16]. In contrast, their role in the evaluation of specimens such as pleural fluid is not clear. Our results indicate that commercial NAA tests have high specificity and positive likelihood ratios. These test properties suggest a potential role for commercial tests in confirming (ruling in) the diagnosis of tuberculous pleuritis. These tests, however, have low and widely varying sensitivities—test properties that make them unhelpful in ruling out TB. Potential explanations for the low sensitivity include a low bacillary load in pleural fluid, or the presence of substances in the pleural fluid that inhibit amplification [65]. Some authors have suggested that pleural fluids should be tested with NAA methods after the specimens are adequately pre-treated to remove inhibitors [65]. All commercial kits appear to be designed to maximize only specificity. The MTD and LCx kits appear to have higher sensitivity than the Amplicor test. This comparison should be interpreted cautiously because it is based on few studies. Studies that directly compare these commercial tests (head-to-head) within the same study population are required to confirm these observations. The most important finding about in-house PCR is the significant heterogeneity across studies.
Clinical implications

To interpret the summary measures in a clinical context, consider a patient from a high incidence setting (e.g. countries such as Spain or Malaysia) who is estimated to have a 50% probability of TB after clinical evaluation, and is evaluated with either the MTD test (LR+ of 17.4 and LR- of 0.31) or the Amplicor test (LR+ of 52.8 and LR- of 0.59). A LR+ of 17.4 for the MTD test suggests that patients with tuberculous pleuritis have a 17-fold higher chance of being MTD test positive as compared to patients without TB. If the MTD test were positive, the likelihood that this patient has TB increases from 50% to 95%, a probability that is sufficiently high to justify initiation of anti-tuberculosis treatment. A positive Amplicor test will raise the probability of TB from 50% to 97%. In contrast, if the MTD test result were negative, there is still a 24% chance that this patient has TB, probably not sufficiently low to rule out TB with confidence. In case of the Amplicor test, a negative test will reduce the probability from 50% to 40%, again not low enough to rule out TB.

Consider another patient from a low incidence setting (e.g. countries such as the USA), where the baseline probability of TB is low (e.g. 5%). If MTD test were positive, the likelihood that this patient has TB increases from 5% to 48%, a probability that justifies further investigation. A positive Amplicor test will raise the probability of TB from 5% to 75%. If the MTD® test result were negative, the baseline probability changes from 5% to 2%, a negligible shift that is unlikely to be helpful in clinical decision-making. In case of the Amplicor test, the probability changes from 5% to 4%. These examples illustrate the impact of the baseline prevalence (pre-test probability) on predictive values of the tests.
The accuracy of in-house PCR was heterogeneous across studies, and thus meaningful summary measures of accuracy could not be determined. The clinical implications, therefore, will depend on the setting. Institutions that use in-house PCR will have to rely on local data to decide on its accuracy and clinical applicability. In general, PCR for tuberculosis is known to have poor inter-laboratory reproducibility [66].

In addition to the effect of diagnostic thresholds seen in the SROC plot, we identified two factors that were associated with heterogeneity among in-house tests: use of a case-control design, and use of the IS6110 target sequence. Case-control studies sample patients from the extreme ends of the clinical spectrum (an ideal, “extreme contrast” setting). If the sensitivity of a test is evaluated in seriously diseased subjects, and specificity in healthy individuals, both measures will overestimate the true diagnostic accuracy [67]. Empiric research suggests that case-control studies overestimate the diagnostic odds ratio by a factor of 3, when compared to cross-sectional studies [68]. Future studies of NAA tests could reduce this bias by avoiding the case-control design and recruiting consecutive series of patients in whom the test is clinically indicated (a realistic, “clinical practice” setting). The IS6110 target sequence is considered the “gold standard” for \textit{M. tuberculosis} fingerprinting [69]. Because this target is specific to the \textit{M. tuberculosis} complex, and because it is usually present as multiple copies in the genome, PCR tests using this target might be more sensitive. Further research is underway to confirm this finding, in a larger meta-analysis of in-house PCR in the diagnosis of pulmonary tuberculosis.
Previous meta-analyses of NAA test accuracy

Our data are consistent with the results of two previous meta-analyses on the accuracy of NAA tests. Sarmiento and colleagues summarized the accuracy of PCR in the diagnosis of smear-negative pulmonary TB [70]. Their meta-analysis of 50 studies showed that both sensitivity and specificity estimates were heterogeneous. They concluded, PCR is not consistently accurate enough to be routinely recommended for the diagnosis of smear-negative TB. Our previous meta-analysis of 49 studies summarized the accuracy of NAA tests in the diagnosis of tuberculous meningitis [71]. Commercial tests were found to have high overall specificity (0.98) and low sensitivity (0.56). The accuracy of in-house PCR was not determined because of heterogeneity in study results.

Limitations of the review

Our review has limitations. Our analysis lacks data on the incremental gain of NAA tests over and above the diagnostic performance achieved by using only conventional methods or other rapid tests like ADA and IFN-γ. The primary studies in our review did not report such data. Also, few studies in our review directly compared the NAA test against tests such as ADA and IFN-γ [45,51,59]. Only one study [59] directly compared the three tests in the same population, and showed that ADA, IFN-γ and PCR were 88%, 86%, and 74% sensitive, respectively, and 86%, 97%, and 90% specific, respectively, for culture or biopsy-confirmed pleural TB. Since we did not include tests such as ADA and IFN-γ in our literature searches, our review cannot identify the most accurate test. Also, publication bias was a concern with the in-house tests. Exclusion of studies published in languages other than English and Spanish could have contributed to this potential bias.
CONCLUSIONS

In summary, our data suggest a potentially useful role for commercial NAA tests in confirming a diagnosis of tuberculous pleuritis. However, commercial kits have low and varying sensitivities, and therefore should not be used for excluding a diagnosis of tuberculous pleuritis. NAA test results, therefore, cannot replace conventional tests; they need to be interpreted in parallel with clinical findings and results of conventional tests. The accuracy of in-house PCR tests is poorly defined because of heterogeneity in study results. Clinically useful summary measures cannot be estimated for in-house PCR tests.
LIST OF ABBREVIATIONS

ADA: Adenosine Deaminase
DOR: diagnostic odds ratio
FPR: false positive rate
HIV: human immunodeficiency virus
IFN-γ: Interferon-gamma
LR+: positive likelihood ratio
LR–: negative likelihood ratio
NAA: nucleic acid amplification
PCR: polymerase chain reaction
ROC: receiver operating characteristic curve
SROC: summary receiver operating characteristic curve
TB: tuberculosis
TPR: true positive rate
COMPETING INTERESTS

Competing interests: none declared.
AUTHOR CONTRIBUTIONS

Study concept and design: MP, LLF, LWR, JMC.

Acquisition of data: MP, LLF.

Analysis and interpretation of data: MP, AH, LWR, JMC.

Drafting of manuscript: MP, JMC

Critical revision of the manuscript for important intellectual content: all authors

All authors read and approved the final manuscript.
ACKNOWLEDGMENTS

Presented as a poster at the California Tuberculosis Controllers Association Spring 2003 Conference, Oakland, USA, April 24-25, 2003. MP acknowledges the support of the National Institutes of Health, Fogarty AIDS International Training Program (1-D43-TW00003-15). LLF is supported by the Fogarty Tuberculosis supplementary grant (TW00905-S1) and the Consejo Nacional de Ciencia y Tecnologia (CONACyT), National Council of Science and Technology, Mexico (scholarship number 129617). We are grateful to the authors of the primary studies who sent additional data, and representatives of the commercial test manufacturers for sending us lists of studies.
REFERENCES


FIGURE LEGENDS

Figure 1: Study selection process

Figure 2: Forest plots of estimates of sensitivity and specificity in studies with commercial and in-house tests.

The point estimates of sensitivity and specificity from each study are shown as solid circles. Error bars are 95% confidence intervals. Numbers indicate the studies cited in the bibliography. Pooled estimates are summary random effects estimates with 95% confidence intervals.

Figure 3: Summary Receiver Operative Characteristic curves for commercial and in-house tests

Each solid circle represents each study in the meta-analysis. The size of each study is indicated by the size of the solid circle. The weighted (dark line) and unweighted (thin line) regression SROC curves summarize the overall diagnostic accuracy.

Figure 4: Funnel plot for evaluation of publication bias in studies with in-house tests

The funnel graph plots the log of the diagnostic odds ratio (DOR) against the standard error of the log of the DOR (an indicator of sample size). Each open circle represents each study in the meta-analysis. The line in the center indicates the summary DOR. In the absence of publication bias, the DOR estimates from smaller studies are expected to be scattered above and below the summary estimate, producing a triangular or funnel shape. The funnel plots appear asymmetric—smaller studies with low DOR estimates are missing—indicating a potential for publication bias. The Egger test for publication bias was statistically significant (p=0.002) in the in-house test subgroup.
Table 1. Description of included studies—commercial tests

<table>
<thead>
<tr>
<th>Country, year</th>
<th>Independent comparison study design</th>
<th>Double or single blind</th>
<th>Consecutive or random sample</th>
<th>Prospective or cross-sectional study design</th>
<th>Type of testb (target sequence)</th>
<th>Reference standardc</th>
<th>Study qualityd</th>
<th>No. of pleural specimens or subjects with TB/No. without TB</th>
<th>Sensitivity 95% CI</th>
<th>Specificity 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AMPLICOR® Test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA, 28 1996</td>
<td>Yes</td>
<td>Yes</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Amplicor (16s rRNA)</td>
<td>Culture Low</td>
<td>Yes</td>
<td>(3/89) (0.67 (0.13, 0.98))</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Japan, 42 2000</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Amplicor (16s rRNA)</td>
<td>Culture biopsy Medium</td>
<td>No</td>
<td>(33/42) (0.27 (0.14, 0.46))</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>Germany, 54 1998</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Amplicor (16s rRNA)</td>
<td>Culture High</td>
<td>Yes</td>
<td>(3/66) (0.67 (0.13, 0.98))</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>USA, 55 1998</td>
<td>Yes</td>
<td>Yes</td>
<td>Unknown</td>
<td>Yes</td>
<td>Amplicor (16s rRNA)</td>
<td>Culture Medium</td>
<td>Yes</td>
<td>(8/367) (0.50 (0.18, 0.82))</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td><strong>LCx® Test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argentina, 44 2001</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>LCx (38 kDa)</td>
<td>Culture clinical High</td>
<td>Yes</td>
<td>(16/186) (0.69 (0.42, 0.88))</td>
<td>0.96</td>
<td>(0.92, 0.98)</td>
</tr>
<tr>
<td>Spain, 46 1998</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>LCx (38 kDa)</td>
<td>Culture biopsy High</td>
<td>Yes</td>
<td>(15/171) (0.87 (0.59, 0.98))</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>Finland, 52 2001</td>
<td>Partially</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>LCx (38 kDa)</td>
<td>Culture Medium</td>
<td>No</td>
<td>(5/142) (0.40 (0.07, 0.83))</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Spain, 57 1999</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>LCx (38 kDa)</td>
<td>Culture Medium</td>
<td>Yes</td>
<td>(3/12) (1.00 (0.29, 1.00))</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Country, year</td>
<td>Independent comparison</td>
<td>Cross-sectional study design</td>
<td>Double or single blind</td>
<td>Consecutive or random sample</td>
<td>Prospective data collection</td>
<td>Type of test(^b) (target sequence)</td>
<td>Reference standard(^c)</td>
<td>Study quality(^d)</td>
<td>No. of pleural specimens or subjects with TB/No. without TB</td>
<td>Sensitivity (^e) (95% CI)</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------</td>
<td>------------------------------</td>
<td>------------------------</td>
<td>-----------------------------</td>
<td>------------------------------</td>
<td>---------------------------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>---------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Spain, 2001</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>MTD-2 (rRNA)</td>
<td>Culture/clinical</td>
<td>Medium</td>
<td>5/96</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>Germany, 1996</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Partly</td>
<td>MTD-1 (rRNA)</td>
<td>Culture/clinical</td>
<td>Medium</td>
<td>3/32</td>
<td>1.00</td>
</tr>
<tr>
<td>Spain, 1997</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>MTD-1 (rRNA)</td>
<td>Culture</td>
<td>High</td>
<td>13/36</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Spain, 1997</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>MTD-1 (rRNA)</td>
<td>Culture</td>
<td>High</td>
<td>13/28</td>
<td>1.00</td>
</tr>
<tr>
<td>Switzerland</td>
<td>Yes</td>
<td>Yes</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>MTD-1 (rRNA)</td>
<td>Culture</td>
<td>Low</td>
<td>4/61</td>
<td>0.50</td>
</tr>
<tr>
<td>Netherlands, 1995</td>
<td>Yes</td>
<td>Yes</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>MTD-1 (rRNA)</td>
<td>Culture</td>
<td>Medium</td>
<td>5/56</td>
<td>0.20</td>
</tr>
</tbody>
</table>

\(^a\) Incorporates additional information provided by authors of the primary studies

\(^b\) Type of test: Amplicor (Roche Molecular Systems, Branchburg, NJ); MTD (Gen-Probe Inc, San Diego, CA); LCx (Abbott Laboratories, Abbott Park, IL)

\(^c\) Reference standard: culture refers to culture of pleural fluid/sputum/biopsy; biopsy refers to pleural biopsy; clinical refers to data such as signs and symptoms, response to therapy, and pleural fluid analysis

\(^d\) See text for description of quality assessment; \(^e\) CI: Confidence Interval
Table 2. Description of included studies—In-house PCR tests

<table>
<thead>
<tr>
<th>Country, year</th>
<th>Independent comparison</th>
<th>Cross-sectional study design</th>
<th>Double or single blind</th>
<th>Consecutive or random sample</th>
<th>Prospective data collection</th>
<th>Type of testb (target sequence)</th>
<th>Reference standardc</th>
<th>Study qualityd</th>
<th>No. of pleural specimens or subjects with TB/No. without TB</th>
<th>Sensitivity (95% CI)e</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brazil, 2000</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Partly</td>
<td>PCR (16S rRNA)</td>
<td>Culture/clinical</td>
<td>Medium</td>
<td>28/17</td>
<td>0.89 (0.71, 0.97)</td>
<td>1.00 (0.80, 1.00)</td>
</tr>
<tr>
<td>Spain, 2000</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>PCR (IS6110)</td>
<td>Culture/clinical</td>
<td>Medium</td>
<td>9/16</td>
<td>0.89 (0.51, 0.99)</td>
<td>1.00 (0.79, 1.00)</td>
</tr>
<tr>
<td>China, 1996</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>PCR (IS6110/38 kDa)</td>
<td>Culture/medium</td>
<td>25/76</td>
<td>0.44</td>
<td>0.99 (0.25, 0.65)</td>
<td>0.92 (0.92, 1.00)</td>
</tr>
<tr>
<td>India, 1998</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Unknown</td>
<td>Yes</td>
<td>PCR (MPB64)</td>
<td>Culture/clinical</td>
<td>Medium</td>
<td>28/16</td>
<td>0.61 (0.41, 0.78)</td>
<td>0.94 (0.68, 1.00)</td>
</tr>
<tr>
<td>France, 1992</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>PCR (IS6110)</td>
<td>Culture/biopsy</td>
<td>Low</td>
<td>15/10</td>
<td>0.60 (0.33, 0.82)</td>
<td>1.00 (0.69, 1.00)</td>
</tr>
<tr>
<td>France, 1992</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Partly</td>
<td>PCR (65 kDa)</td>
<td>Culture/biopsy</td>
<td>Low</td>
<td>15/10</td>
<td>0.20 (0.05, 0.48)</td>
<td>1.00 (0.69, 1.00)</td>
</tr>
<tr>
<td>S. Africa, 1992</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>PCR (336 bp repetitive sequence)</td>
<td>Culture/biopsy</td>
<td>Medium</td>
<td>50/31</td>
<td>0.82 (0.68, 0.91)</td>
<td>0.77 (0.59, 0.90)</td>
</tr>
<tr>
<td>S. Africa, 1990</td>
<td>Yes</td>
<td>Yes</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>PCR (336 bp repetitive sequence)</td>
<td>Culture/medium</td>
<td>Medium</td>
<td>11/5</td>
<td>1.00 (0.72, 1.00)</td>
<td>1.00 (0.48, 1.00)</td>
</tr>
</tbody>
</table>
Table 2 (continued). Description of included studies—In-house PCR tests

<table>
<thead>
<tr>
<th>Country, year</th>
<th>Independent comparison</th>
<th>Cross-sectional study design</th>
<th>Double or single blind</th>
<th>Consecutive or random sample</th>
<th>Prospective or retrospective data collection</th>
<th>Type of test (target sequence)</th>
<th>Reference standard</th>
<th>Study quality</th>
<th>No. of pleural specimens or subjects with TB/No. without TB</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>India, 2001</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>PCR (IS6110)</td>
<td>Culture Medium</td>
<td>4/26</td>
<td>0.25</td>
<td>(0.01, 0.78)</td>
<td>0.96</td>
</tr>
<tr>
<td>Chile, 1993</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Unknown</td>
<td>Unknown</td>
<td>PCR (65 kDa)</td>
<td>Culture/ microscopy</td>
<td>15/11</td>
<td>0.93</td>
<td>(0.66, 1.00)</td>
<td>1.00</td>
</tr>
<tr>
<td>Mexico et al, 2001</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>PCR (465 bp sequence)</td>
<td>Culture High</td>
<td>3/12</td>
<td>0.33</td>
<td>(0.02, 0.87)</td>
<td>0.83</td>
</tr>
<tr>
<td>France, 1996</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>PCR (IS6110)</td>
<td>Culture/biopsy Medium</td>
<td>17/25</td>
<td>0.76</td>
<td>(0.50, 0.92)</td>
<td>1.00</td>
</tr>
<tr>
<td>Brazil, 2000</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Partly</td>
<td>PCR (MPB64)</td>
<td>Culture/biopsy High</td>
<td>19/11</td>
<td>0.68</td>
<td>(0.44, 0.86)</td>
<td>0.91</td>
</tr>
<tr>
<td>Japan, 1993</td>
<td>Yes</td>
<td>Yes</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>PCR (38 kDa)</td>
<td>Culture Low</td>
<td>5/105</td>
<td>0.80</td>
<td>(0.30, 0.99)</td>
<td>0.93</td>
</tr>
<tr>
<td>India, 2001</td>
<td>Yes</td>
<td>Yes</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>PCR (150 bp sequence)</td>
<td>Microscopy/biopsy High</td>
<td>20/40</td>
<td>0.70</td>
<td>(0.46, 0.87)</td>
<td>1.00</td>
</tr>
<tr>
<td>Taiwan, 1990</td>
<td>Yes</td>
<td>Yes</td>
<td>Unknown</td>
<td>Unknown</td>
<td>PCR (65 kDa)</td>
<td>Culture Medium</td>
<td>11/37</td>
<td>1.00</td>
<td>(0.72, 1.00)</td>
<td>0.78</td>
<td></td>
</tr>
</tbody>
</table>
Table 2 (continued). Description of included studies—In-house PCR tests\textsuperscript{a}

<table>
<thead>
<tr>
<th>Country, year</th>
<th>Independent comparison</th>
<th>Cross-sectional study design</th>
<th>Double or single blind</th>
<th>Consecutive or random sample</th>
<th>Prospective or retrospective data collection</th>
<th>Type of test\textsuperscript{b} (target sequence)</th>
<th>Reference standard\textsuperscript{c}</th>
<th>Study quality\textsuperscript{d}</th>
<th>No. of pleural specimens or subjects with TB/No. without TB</th>
<th>Sensitivity (95% CI)\textsuperscript{e}</th>
<th>Specificity (95% CI)\textsuperscript{e}</th>
</tr>
</thead>
<tbody>
<tr>
<td>India, 2000</td>
<td>Yes</td>
<td>Yes</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>PCR (IS6110/TRC\textsubscript{4})</td>
<td>Culture/microscopy</td>
<td>Unknown</td>
<td>30/20</td>
<td>1.00</td>
<td>0.70</td>
</tr>
<tr>
<td>Mexico, 2000</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>PCR (IS6110)</td>
<td>Culture/biopsy</td>
<td>High</td>
<td>17/56</td>
<td>0.94</td>
<td>1.00</td>
</tr>
<tr>
<td>Spain, 1995</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>PCR (IS6110)</td>
<td>Culture/biopsy</td>
<td>High</td>
<td>21/86</td>
<td>0.81</td>
<td>0.98</td>
</tr>
<tr>
<td>Thailand, 2000</td>
<td>Yes</td>
<td>Yes</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Yes</td>
<td>PCR (16S rRNA)</td>
<td>Culture/biopsy</td>
<td>High</td>
<td>36/62</td>
<td>0.72</td>
<td>0.53</td>
</tr>
<tr>
<td>Singapore, 1995</td>
<td>Yes</td>
<td>No</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>PCR (IS6110)</td>
<td>Culture/biopsy</td>
<td>Low</td>
<td>10/13</td>
<td>0.70</td>
<td>1.00</td>
</tr>
<tr>
<td>Singapore, 1995</td>
<td>Yes</td>
<td>No</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>PCR (MPB64)</td>
<td>Culture/biopsy</td>
<td>Low</td>
<td>10/13</td>
<td>0.70</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 2 (continued). Description of included studies—In-house PCR tests

<table>
<thead>
<tr>
<th>Country, year</th>
<th>Independent comparison study design</th>
<th>Cross-sectional study or single blind study</th>
<th>Double or single blind</th>
<th>Consecutive or random sample</th>
<th>Prospective data collection</th>
<th>Type of test (target sequence)</th>
<th>Reference standard</th>
<th>Study quality</th>
<th>No. pleural specimens or subjects with TB/No. without TB</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singapore, 56 1997</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>PCR (IS6110)</td>
<td>Culture/clinical</td>
<td>High</td>
<td>16/49</td>
<td>1.00 (0.79, 1.00)</td>
<td>0.88 (0.75, 0.95)</td>
</tr>
<tr>
<td>India, 58 1995</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>PCR (150 bp sequence)</td>
<td>Clinical/biopsy</td>
<td>Medium</td>
<td>38/29</td>
<td>0.63 (0.46, 0.78)</td>
<td>0.93 (0.76, 0.99)</td>
</tr>
<tr>
<td>Colombia, 59 2000</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>PCR (IS6110)</td>
<td>Culture/biopsy</td>
<td>High</td>
<td>42/70</td>
<td>0.74 (0.58, 0.86)</td>
<td>0.90 (0.80, 0.96)</td>
</tr>
<tr>
<td>Spain, 60 1998</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>PCR (IS6110)</td>
<td>Culture/biopsy</td>
<td>High</td>
<td>33/98</td>
<td>0.42 (0.26, 0.61)</td>
<td>0.99 (0.94, 1.00)</td>
</tr>
</tbody>
</table>

* Incorporates additional information provided by authors of the primary studies

* Type of test: PCR: polymerase chain reaction

* Reference standard: culture refers to culture of pleural fluid/sputum/biopsy; biopsy refers to pleural biopsy; clinical refers to data such as signs and symptoms, response to therapy, and pleural fluid analysis

* See text for description of quality assessment

* CI: Confidence Interval
Table 3: Summary accuracy measures for commercial and in-house tests

<table>
<thead>
<tr>
<th>Measure of test accuracy</th>
<th>Pooled summary measure&lt;sup&gt;a&lt;/sup&gt; (95% CI)</th>
<th>Test for heterogeneity&lt;sup&gt;b&lt;/sup&gt; p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Commercial Tests [N = 14]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.62 (0.43, 0.77)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.98 (0.96, 0.98)</td>
<td>0.12</td>
</tr>
<tr>
<td>Positive Likelihood Ratio (LR+)</td>
<td>25.4 (16.2, 40.0)</td>
<td>0.46</td>
</tr>
<tr>
<td>Negative likelihood Ratio (LR–)</td>
<td>0.40 (0.24, 0.67)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diagnostic odds ratio (DOR)</td>
<td>80.9 (34.4, 190.4)</td>
<td>0.09</td>
</tr>
<tr>
<td><strong>Amplicor [N = 4]</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.37 (0.24, 0.52)</td>
<td>0.18</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.99 (0.97, 1.0)</td>
<td>0.12</td>
</tr>
<tr>
<td>Positive Likelihood Ratio (LR+)</td>
<td>52.8 (11.8, 236.2)</td>
<td>0.14</td>
</tr>
<tr>
<td>Negative likelihood Ratio (LR–)</td>
<td>0.59 (0.37, 0.95)</td>
<td>0.20</td>
</tr>
<tr>
<td>Diagnostic odds ratio (DOR)</td>
<td>105.5 (15.1, 733.9)</td>
<td>0.10</td>
</tr>
<tr>
<td><strong>LCx [N = 4]</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.72 (0.54, 0.84)</td>
<td>0.12</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.97 (0.95, 0.98)</td>
<td>0.53</td>
</tr>
<tr>
<td>Positive Likelihood Ratio (LR+)</td>
<td>25.6 (14.1, 46.4)</td>
<td>0.57</td>
</tr>
<tr>
<td>Negative likelihood Ratio (LR–)</td>
<td>0.32 (0.14, 0.68)</td>
<td>0.09</td>
</tr>
<tr>
<td>Diagnostic odds ratio (DOR)</td>
<td>93.2 (35.6, 243.7)</td>
<td>0.38</td>
</tr>
<tr>
<td><strong>MTD [N = 6]</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.77 (0.62, 0.88)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.98 (0.96, 0.99)</td>
<td>0.66</td>
</tr>
<tr>
<td>Positive Likelihood Ratio (LR+)</td>
<td>17.4 (6.1, 49.7)</td>
<td>0.26</td>
</tr>
<tr>
<td>Negative likelihood Ratio (LR–)</td>
<td>0.31 (0.09, 1.03)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diagnostic odds ratio (DOR)</td>
<td>72.9 (9.9, 533.4)</td>
<td>0.03</td>
</tr>
<tr>
<td><strong>In-house Tests [N = 26]</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.71 (0.63, 0.78)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.93 (0.88, 0.96)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Positive Likelihood Ratio (LR+)</td>
<td>9.7 (5.7, 16.6)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Negative likelihood Ratio (LR–)</td>
<td>0.32 (0.23, 0.43)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diagnostic odds ratio (DOR)</td>
<td>42.4 (22.2, 81.1)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<sup>a</sup> Random effects model; <sup>b</sup> Chi-square or Fisher’s exact test for heterogeneity
Table 4: Stratified analyses for the evaluation of heterogeneity in studies with in-house tests

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Summary Diagnostic</th>
<th>Test for heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds Ratio (95% CI)</td>
<td>p value</td>
</tr>
</tbody>
</table>

**Study design**

- Case-control (N = 12 studies) 68.3 (31.8, 146.5) 0.39
- Cross-sectional (N = 14 studies) 28.7 (12.3, 66.9) 0.002

**Target sequence**

- IS6110 (N = 12 studies) 62.4 (32.8, 118.7) 0.42
- Other target sequences (N = 14 studies) 25.5 (10.6, 61.3) 0.006

*aConfidence interval

*bChi-square test for heterogeneity
2520 potentially relevant citations identified from all searches

2086 citations excluded after initial screen of citations

434 English and Spanish articles selected for full-text review

65 articles excluded after full-text screen
  17 No original data
  11 Fingerprinting study
  8 No comparison against reference standard
  7 Case reports
  6 Description of new test methods
  5 Reliability study
  3 Treatment monitoring study
  8 Other reasons

369 articles met eligibility criteria (for all forms of tuberculosis)

71 articles excluded
  56 No sensitivity and specificity data for pleural specimens
  12 Less than 10 pleural specimens
  3 Samples tested were pleural tissue/cytology specimens

109 articles reported inclusion of pleural specimens (fluid and/or pleural biopsy)

38 articles included in the meta-analysis
  34 English
  4 Spanish
(A) Commercial Tests

(B) In-house Tests
Figure 3
Begg's funnel plot with pseudo 95% confidence limits

Figure 4