# Diagnostic Value of Interferon-γ in Tuberculous Pleurisy*

## A Metaanalysis

*Jing Jiang, MD; Huan-Zhong Shi, MD, PhD; Qiu-Li Liang, MD; Shou-Ming Qin, MD; and Xue-Jun Qin, MD*

***Background:*** **Conventional tests are not always helpful in making a diagnosis of tuberculous pleurisy. Many studies have investigated the usefulness of interferon (IFN)-γ measurements in pleural fluid for the early diagnosis of tuberculous pleurisy. We conducted a metaanalysis to determine the accuracy of IFN-γ measurements in the diagnosis of tuberculous pleurisy.**

***Methods:*** **After a systematic review of English-language studies, sensitivity, specificity, and other measures of accuracy of IFN-γ concentrations in the diagnosis of pleural effusion were pooled using random-effects models. Summary receiver operating characteristic curves were used to summarize overall test performance.**

***Results:*** **Twenty-two studies met our inclusion criteria. The summary estimates for IFN-γ in the diagnosis of tuberculous pleurisy in the studies included were as follows: sensitivity, 0.89 (95% confidence interval [CI], 0.87 to 0.91); specificity, 0.97 (95% CI, 0.96 to 0.98); positive likelihood ratio, 23.45 (95% CI, 17.31 to 31.78); negative likelihood ratio, 0.11 (95% CI, 0.07 to 0.16); and diagnostic odds ratio, 272.7 (95% CI, 147.5 to 504.2).**

***Conclusions:*** **IFN-γ determination is a sensitive and specific test for the diagnosis of tuberculous pleurisy. The measurement of IFN-γ levels in pleural effusions is thus likely to be a useful tool for diagnosing tuberculous pleurisy. The results of IFN-γ assays should be interpreted in parallel with clinical findings and the results of conventional tests.** ***(CHEST 2007; 131:1133–1141)***

**Key words:** interferon; pleural effusion; tuberculosis

**Abbreviations:** AUC = area under the curve; CI = confidence interval; DOR = diagnostic odds ratio; IFN = interferon; NLR = negative likelihood ratio; PLR = positive likelihood ratio; QUADAS = quality assessment for studies of diagnostic accuracy; RDOR = relative diagnostic odds ratio; ROC = receiver operating characteristic; SROC = summary receiver operating characteristic; STARD = standards for reporting diagnostic accuracy; TPE = tuberculous pleural effusion

Tuberculosis is the leading cause of death from a curable infectious disease. In China, the prevalence of active pulmonary tuberculosis in 2000 was 367 per 100,000 population, the prevalence of smear positive pulmonary tuberculosis was 122 per 100,000 population, and the prevalence of bacteriological positive pulmonary tuberculosis was 160 per 100,000 population.[1] On the basis of results of surveys of the prevalence of infection and disease, assessments of the effectiveness of surveillance systems, and death registrations, there were an estimated 8.9 million new cases of tuberculosis in 2004, fewer than half of which were reported to public-health authorities and World Health Organization.[2] Tuberculous pleural effusion (TPE) is caused by a severe delayed-type hypersensitivity reaction in response to the rupture of a subpleural focus of *Mycobacterium tuberculosis* infection. Although TPE occurs in about 10% of untreated individuals who test positive by the tuberculin test, it may also develop as a complication of primary pulmonary tuberculosis.[3] Actually, tuberculosis is the major cause of pleural effusions in areas of high tuberculosis prevalence, and TPE usually manifests as lymphocytic exudative effusion.[4] Many studies have investigated the usefulness of interferon (IFN)-γ in pleural fluid for the early diagnosis of tuberculous pleural effusion (TPE), and an early metaanalysis has shown that the value of pleural

IFN-γ measurements for the diagnosis of TPE was reasonably good[5]; however, likelihood ratios including both positive likelihood ratio (PLR) and negative likelihood ratio (NLR), and diagnostic odds ratio (DOR) have not been evaluated in the metaanalysis. In that metaanalysis, 13 early related studies were included. Since that time, additional clinical studies determining the concentrations of IFN-γ have been reported. It is well defined that accuracy is the degree of conformity of a measured or calculated value to its actual or specified value. Although the accuracy of IFN-γ detection for the diagnosis of TPE has been extensively studied, the exact role of these detections remains controversial. We performed the present metaanalysis to establish the overall accuracy of IFN-γ measurement for the diagnosis of TPE.

## Materials and Methods

### *Search Strategy and Study Selection*

Since the present study was a metaanalysis that was based on published articles, we did not include the consents of patients and the approval of internal review boards. We searched the following electronic databases: Medline (1980 to 2006); Embase (1980 to 2006); Web of Science (1990 to 2006); BIOSIS (1993 to 2006); and LILACS (1980 to 2006). We also reviewed the Cochrane Library to find relevant articles. All searches were up to date as of October 2006. The search terms used were "tuberculosis," "*Mycobacterium tuberculosis*," "pleurisy/pleuritis," "pleural effusion/pleural fluid," "interferon/IFN," "sensitivity and specificity," and "accuracy." We contacted experts in the specialty, and searched the reference lists from primary and review articles. Although no language restrictions were imposed initially, for the full-text review and final analysis our resources only permitted the review of articles published in the English language. Conference abstracts and letters to the journal editors were excluded because of the limited data presented in them.

A study was included in the metaanalysis when it provided both the sensitivity (true-positive rate) and the specificity (false-positive rate) of IFN-γ for the diagnosis of TPE, or when it provided IFN-γ values in a dot-plot form, allowing test results to be extracted for individual study subjects. The studies including at least 10 TPE specimens were selected for inclusion in the study, since very small studies may be vulnerable to selection bias. Two reviewers (J.J. and H.Z.S.) independently judged study eligibility while screening the citations. Disagreements were resolved by consensus.

### *Data Extraction and Quality Assessment*

The final set of English language articles was assessed independently by two reviewers (J.J. and H.Z.S.). The reviewers were blinded to publication details, and disagreements were resolved by consensus. Data retrieved from the reports included participant characteristics, test methods, sensitivity and specificity data, cutoff values, publication year, and methodological quality. Where pleural IFN-γ values were provided in dot plots, scalar grids were placed over the plots, and the values were measured and used to produce a receiver operating characteristic (ROC) curve for each study (SPSS; Chicago, IL). The numbers of true-positive, false-positive, false-negative, and true-negative results are displayed for each study in Table 1.

We assessed the methodological quality of the studies using guidelines published by the standards for reporting diagnostic accuracy (STARD) initiative[6] (maximum score, 25) [*ie*, guidelines that aim to improve the quality of reporting in diagnostic studies] and the quality assessment for studies of diagnostic accuracy (QUADAS) tool[7] (maximum score, 14) [*ie*, appraisal by use of empirical evidence, expert opinion, and formal consensus to assess the quality of primary studies of diagnostic accuracy]. In addition, for each study the following characteristics of study design were also retrieved: (1) cross-sectional design (vs case-control design); (2) consecutive or random sampling of patients; (3) blinded (single or double) interpretation of determination and reference standard results; and (4) prospective data collection. If no data on the above criteria were reported in the primary studies, we requested the information from the authors. If the authors did not respond to our letters, the "unknown" items were treated as "no."

### *Statistical Analysis*

We used standard methods recommended for metaanalyses of diagnostic test evaluations.[8] Analyses were performed using several statistical software programs (Stata, version 8.2; Stata Corporation; College Station, TX; Meta-Test, version 0.6; New England Medical Center; Boston, MA; and Meta-DiSc for Windows; XI Cochrane Colloquium; Barcelona, Spain). We computed the following measures of test accuracy for each study: sensitivity; specificity; PLR; NLR; and DOR.

The analysis was based on a summary ROC (SROC) curve.[8,9] The sensitivity and specificity for the single test threshold identified for each study were used to plot an SROC curve.[9,10] A random-effects model was used to calculate the average sensitivity, specificity, and the other measures across studies.[11,12]

The term *heterogeneity* when used in relation to metaanalyses refers to the degree of variability in results across studies. We used the $\chi^2$ and Fisher exact tests to detect statistically significant

*From the Institute of Respiratory Diseases, First Affiliated Hospital, Guangxi Medical University, Nanning, Guangxi, People's Republic of China.
Dr. Shi designed the study, searched the databases, extracted the data, analyzed the results, and wrote the manuscript. Dr. Jiang helped with study design, searching the databases, and writing and revising the manuscript. Dr. Liang formulated the research question, and helped with database searches and analysis. Drs. S.-M. Qin and X.-J. Qin helped to design the data abstraction form and served as second reviewers in extracting the data. All authors read and approved the final manuscript.
The authors have reported to the ACCP that no significant conflicts of interest exist with any companies/organizations whose products or services may be discussed in this article.
This study was supported in part by research grant 30660064 from National Natural Science Foundation of China, in part by Program NCET-04–0835 for New Century Excellent Talents in Chinese Universities, and in part by research grant 0639044 from the Natural Science Foundation of Guangxi Zhuang Autonomous Zone, People's Republic of China.
Manuscript received September 18, 2006; revision accepted November 22, 2006.
Reproduction of this article is prohibited without written permission from the American College of Chest Physicians (www.chestjournal.org/misc/reprints.shtml).
*Correspondence to: Huan-Zhong Shi, MD, Institute of Respiratory Diseases, First Affiliated Hospital, Guangxi Medical University, Nanning 530021, Guangxi, People's Republic of China; e-mail: hzshi@vip.tom.com*
**DOI: 10.1378/chest.06-2273**

**Table 1—*Summary of Included Studies**

| Study/Year | Patients, No. | Assay Method | Cutoff | Test Results | | | | Quality Score | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | TP | FP | FN | TN | STARD | QUADAS |
| Ribera et al[16]/1988 | 80 | RIA | 2 IU/mL | 30 | 0 | 0 | 50 | 13 | 12 |
| Hsu et al[17]/1989 | 39 | ELISA | 10 IU/mL | 18 | 0 | 1 | 20 | 11 | 9 |
| Shimokata et al[18]/1991 | 40 | RIA | Unknown | 20 | 1 | 0 | 19 | 11 | 12 |
| Maeda et al et al[19]/1993 | 21 | ELISA | 300 pg/mL | 9 | 1 | 5 | 6 | 8 | 8 |
| Valdes et al[20]/1993 | 145 | ELISA | 140 pg/mL | 33 | 9 | 2 | 101 | 11 | 6 |
| Aoki et al[21]/1994 | 39 | ELISA | 0.3 IU/mL | 11 | 0 | 0 | 28 | 13 | 9 |
| Soderblom et al[22]/1996 | 102 | ELISA | 1.5 pg/mL | 43 | 0 | 11 | 48 | 14 | 10 |
| Kim et al[23]/1997 | 70 | RIA | 9.1 IU/mL | 29 | 2 | 10 | 29 | 15 | 7 |
| Ogawa et al[24]/1997 | 50 | ELISA | 5 IU/mL | 17 | 0 | 1 | 32 | 14 | 12 |
| Wongtim et al[25]/1999 | 66 | ELISA | 240 pg/mL | 37 | 1 | 2 | 26 | 16 | 10 |
| Villegas et al[26]/2000 | 137 | ELISA | 6 IU/mL | 45 | 2 | 13 | 77 | 13 | 8 |
| Yamada et al[27]/2001 | 70 | RIA | 3.1 IU/mL | 20 | 0 | 1 | 49 | 11 | 9 |
| Aoe et al[28]/2003 | 46 | ELISA | 5.7 IU/mL | 10 | 0 | 0 | 36 | 10 | 9 |
| Villena et al[29]/2003 | 595 | RIA | 3.7 IU/mL | 80 | 12 | 2 | 501 | 14 | 12 |
| Wong et al[30]/2003 | 66 | ELISA | 60 pg/mL | 32 | 0 | 0 | 34 | 16 | 13 |
| Poyraz et al[31]/2004 | 45 | ELISA | 12 pg/mL | 13 | 1 | 2 | 29 | 13 | 8 |
| Sharma and Banga[32]/2004 | 101 | ELISA | 138 pg/mL | 58 | 1 | 6 | 36 | 16 | 12 |
| El-Ansary and Radwan[33]/2005 | 39 | ELISA | 3.1 IU/mL | 14 | 0 | 1 | 24 | 10 | 9 |
| Gao and Tian[34]/2005 | 190 | ELISA | 61.7 pg/mL | 119 | 2 | 22 | 47 | 13 | 12 |
| Okamoto et al[35]/2005 | 43 | ELISA | 99.3 pg/mL | 10 | 1 | 1 | 31 | 10 | 9 |
| Sharma and Banga[36]/2005 | 52 | ELISA | 167.5 pg/mL | 34 | 0 | 1 | 17 | 16 | 10 |
| Morimoto et al[37]/2006 | 65 | ELISA | 248 pg/mL | 16 | 3 | 3 | 43 | 11 | 10 |

*ELISA = enzyme-linked immunosorbent assay; RIA = radioimmunoassay; TP = true-positive; FP = false-positive; FN = false-negative; TN = true-negative.

heterogeneity. To assess the effects of STARD and QUADAS scores on the diagnostic ability of IFN-γ, we included them as covariates in univariate metaregression analysis (inverse variance weighted). We also analyzed the effects of other covariates on DOR (*ie*, cross-sectional design, consecutive or random sampling of patients, single or double interpretation of determination and reference standard results, and prospective data collection). The relative DOR (RDOR) was calculated according to standard methods to analyze the change in diagnostic precision in the study per unit increase in the covariate.[13,14] Since publication bias is of concern for metaanalyses of diagnostic studies, we tested for the potential presence of this bias using funnel plots and the Egger test.[15]

## Results

After independent review, 34 publications dealing with pleural IFN-γ concentrations for the diagnosis of TPE were considered to be eligible for inclusion in the analysis.[16–49] Of these publications, two studies[38,39] were excluded because IFN-γ concentration was determined only in TPE patients, two studies[40,41] were excluded because they recruited < 10 patients with confirmed TPE, four studies[42–45] were excluded because they did not allow the calculation of sensitivity or specificity, and four studies[46–49] were excluded because they included groups of patients that had been described elsewhere. The data in the study by Villena et al[46] were reported in another study by Villena et al[29] (V. Villena, MD; personal communication; July 2006). The data in the study by Sharma et al[47] were reported in the study by Sharma and Banga[32], and those in the studies by Hiraki and colleagues[48,49] were reported in the study by Aoe et al.[28] Subsequently, 22 studies[16–37] including 782 patients with TPE and 1,319 non-TPE patients were available for analysis, and the clinical characteristics of these studies, along with QUADAS scores, are outlined in Table 1.

### *Quality of Reporting and Study Characteristics*

The average interrater agreement between the two reviewers for items in the quality checklist was 0.85. The average sample size of the included studies was 96 (range, 21 to 595). In one study,[34] patients with TPE received diagnoses based on clinical presentation, pleural fluid analysis, radiology findings, and the responsiveness of the patient to antituberculous chemotherapy. In two studies,[22,26] a small proportion of the patients studied received diagnoses according to the clinical presentation, pleural fluid analysis, radiology findings, and the responsiveness of the patient to antituberculous chemotherapy, but the diagnosis of pleural tuberculosis was confirmed in most of the TPE patients based on the conventional "gold standard," which is a smear or culture that is positive for *M tuberculosis* taken from pleural fluid and/or histology showing a caseating granu-

loma. In the remaining 19 studies,[16–21,23–25,27–33,35–37] the diagnoses of all patients with TPE were made based on a smear or culture that was positive for *M tuberculosis* that had been taken from pleural fluid and/or histology showing a caseating granuloma. Our initial data were affected by the poor quality of reporting in the primary studies. To overcome this problem, we contacted all authors of the 22 studies included by air mail as well as e-mail when e-mail addresses were available. Seventeen authors responded who could provide additional data for 18 studies.[16–18,20–27,29–32,34,36,37] As shown in Table 2, in 10 of 22 studies (45.5%), the study was cross-sectional design. In 13 studies (59.1%), the samples were collected from consecutive patients. Eleven studies (50.0%) reported blinded interpretation of the IFN-γ assay independent of the reference standard. Fourteen studies (63.6%) that reported that the study design was prospective can be identified from Table 2.

### *Diagnostic Accuracy*

Figure 1 shows the forest plot of sensitivity and specificity for 22 IFN-γ assays in the diagnosis of TPE. The sensitivity ranged from 0.64 to 1.00 (mean, 0.89; 95% confidence interval [CI], 0.87 to 0.91), while specificity ranged from 0.86 to 1.00 (mean, 0.97; 95% CI, 0.96 to 0.98). We also noted that PLR was 23.45 (95% CI, 17.31 to 31.78), NLR was 0.11 (95% CI, 0.07 to 0.16), and DOR was 272.7 (95% CI, 147.5 to 504.2). $\chi^2$ values of sensitivity, specificity, PLR, NLR, and DOR were 67.54 ($p < 0.001$), 32.66 ($p = 0.050$), 21.21 ($p = 0.446$), 62.36 ($p < 0.001$), and 30.04 ($p = 0.091$), respectively, indicating a significant heterogeneity for sensitivity and NLR between studies.

Unlike a traditional ROC plot that explores the effect of varying thresholds (*ie*, cut points for determining test positives) on sensitivity and specificity in a single study, each data point in the SROC plot represents a separate study. The SROC curve presents a global summary of test performance, and shows the tradeoff between sensitivity and specificity. A graph of the SROC curve for the IFN-γ determination showing true-positive rates vs false-positive rates from individual studies is shown in Figure 2. As a global measure of test efficacy we used the Q-value, the intersection point of the SROC curve with a diagonal line from the left upper corner to the right lower corner of the ROC space, which corresponds to the highest common value of sensitivity and specificity for the test. This point does not indicate the only or even the best combination of sensitivity and specificity for a particular clinical setting but represents an overall measure of the discriminatory power of a test. Our data showed that the SROC curve is positioned near the desirable upper left corner of the SROC curve, and that the

**Table 2—*Characteristics of Included Studies**

| Study/Year | TB Patients, No./Non-TB Subjects, No. | Reference Standard | Cross-Sectional Design | Consecutive or Random | Blinded Design | Prospective |
|---|---|---|---|---|---|---|
| Ribera et al[16]/1988 | 30/50 | Bac/His | Yes | Yes | Yes | Yes |
| Hsu et al[17]1989 | 19/20 | Bac/His | Unknown | Yes | Yes | Yes |
| Shimokata et al[18]/1991 | 20/20 | Bac/His | Yes | Yes | Yes | Yes |
| Maeda et al[19]/1993 | 14/7 | Bac/His | Unknown | Unknown | Unknown | Unknown |
| Valdes et al[20]/1993 | 35/110 | Bac/His | Yes | No | No | Yes |
| Aoki et al[21]/1994 | 11/28 | Bac/His | No | Yes | No | Yes |
| Soderblom et al[22]/1996 | 54/48 | Bac/His or Clin | Yes | No | Yes | Yes |
| Kim et al[23]/1997 | 39/31 | Bac/His | No | No | No | No |
| Ogawa et al[24]/1997 | 18/32 | Bac/His | Unknown | Yes | Yes | Yes |
| Wongtim et al[25]/1999 | 39/27 | Bac/His | Yes | Yes | Yes | Yes |
| Villegas et al[26]/2000 | 58/79 | Bac/His or Clin | Yes | Yes | Yes | No |
| Yamada et al[27]/2001 | 21/49 | Bac/His | No | Yes | No | No |
| Aoe et al[28]/2003 | 10/36 | Bac/His | Unknown | Unknown | Unknown | Unknown |
| Villena et al[29]/2003 | 82/513 | Bac/His | Yes | Yes | Yes | Yes |
| Wong et al[30]/2003 | 32/34 | Bac/His | Yes | Yes | Yes | Yes |
| Poyraz et al[31]/2004 | 15/30 | Bac/His | No | No | No | Yes |
| Sharma and Banga[32]/2004 | 64/37 | Bac/His | No | Yes | Yes | Yes |
| El-Ansary and Radwan[33]/2005 | 15/24 | Bac/His | Unknown | Unknown | Unknown | Unknown |
| Gao and Tian[34]/2005 | 141/49 | Clin | Yes | Yes | Yes | Yes |
| Okamoto et al[35]/2005 | 11/32 | Bac/His | Unknown | Unknown | Unknown | Unknown |
| Sharma and Banga[36]/2005 | 35/17 | Bac/His | No | No | Yes | Yes |
| Morimoto et al[37]/2006 | 19/46 | Bac/His | Yes | Yes | No | No |

*TB = tuberculosis; Bac = bacteriology; His = histology; Clin = clinical course.

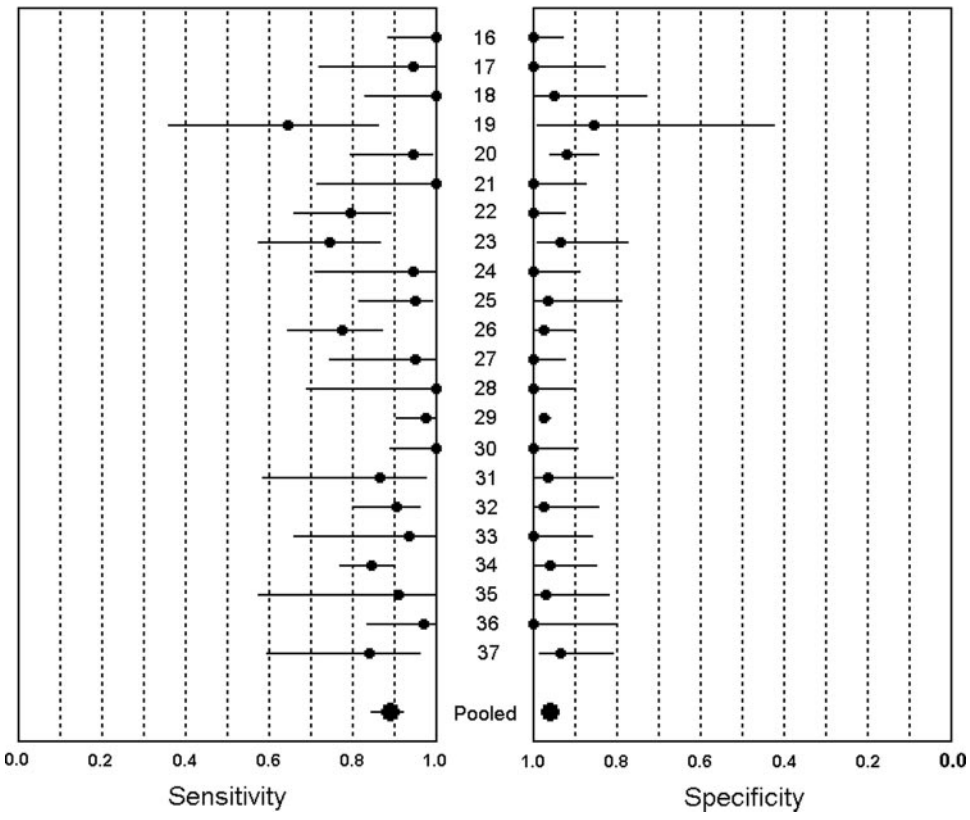


FIGURE 1. Forest plot of estimates of sensitivity and specificity for IFN-γ assays in the diagnosis of TPE. ● = point estimates of sensitivity and specificity from each study; error bars = 95% CIs; numbers = reference numbers of studies cited in the reference list. Pooled estimates for IFN-γ assay were as follows: sensitivity, 0.89 (95% CI, 0.87 to 0.91); specificity, 0.97 (95% CI, 0.96 to 0.98).

maximum joint sensitivity and specificity (*ie*, the Q value) was 0.95; while the area under the curve (AUC) was 0.99 (weighted AUC, 0.98), indicating a high level of overall accuracy.

### *Multiple Regression Analysis and Publication Bias*

By use of the STARD guidelines,[6] a quality score for every study was compiled on the basis of title and introduction, methods, results, and discussion (Table 1). Quality scoring was also done by use of QUADAS,[7] in which a score of 1 was given when a criterion was fulfilled, 0 if a criterion was unclear, and –1 if the criterion was not achieved (Table 1). These scores were used in the metaregression analysis to assess the effect of study quality on the RDOR of IFN-γ in the diagnosis of TPE. As shown in Table 3, studies with higher quality (STARD score, $\geq 13$; QUADAS score, $\geq 10$) produced RDOR values that were not significantly higher than those studies with lower quality. We also noted that differences for studies with or without blinded, cross-sectional, consecutive/random, and prospective designs did not reach statistical significance, indicating that the study design did not substantially affect the diagnostic accuracy.

The evaluation of publication bias showed that the Egger test was significant ($p = 0.023$). The funnel plots for publication bias (Fig 3) also show some asymmetry. These results indicate a potential for publication bias.

## DISCUSSION

Making a differential diagnosis between TPE and non-TPE is a critical clinical problem, and the conventional methods, such as the direct examination of pleural fluid by Ziehl-Neelsen staining, culture of the pleural fluid, and pleural biopsy, are not always helpful in making the diagnosis since they have limitations. Findings of microscopy of the pleural fluid is rarely positive ($< 5\%$).[50–52] Culture of pleural fluid has low sensitivity (24 to 58%), and several weeks are required to grow *M tuberculosis*.[51,53] Biopsy of pleural tissue and culture of biopsy material are widely held to be the best methods of confirming the diagnosis.[50,52] Although not perfect, culture and/or biopsy, therefore, are widely considered to be the standard of diagnosis.[50] However, pleural biopsy is invasive, operator-dependent, and technically difficult (particularly in children).[54]

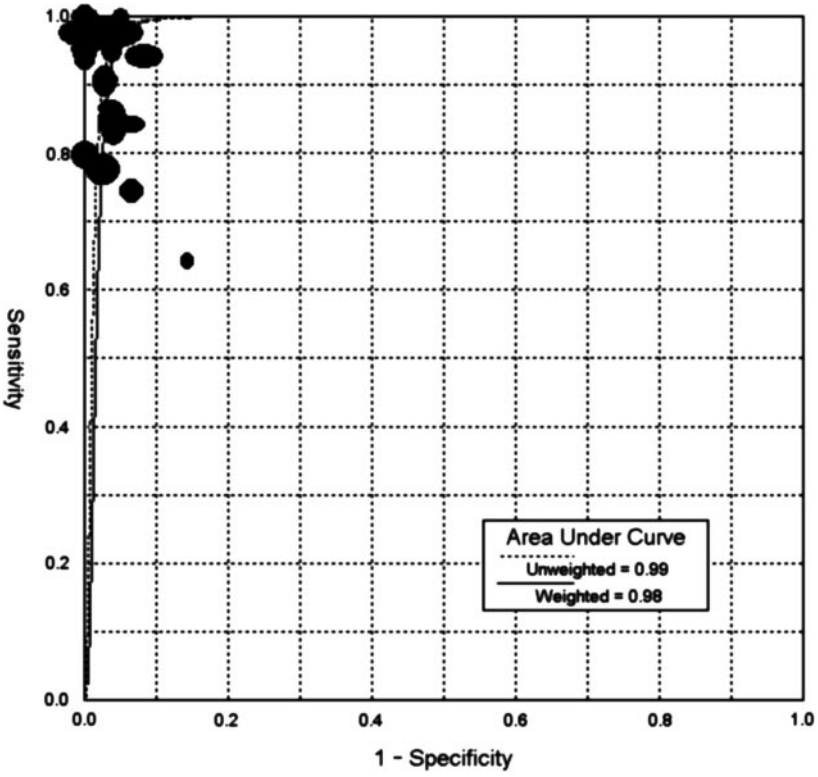


FIGURE 2. SROC curves for IFN-γ assays. ● = each study in the metaanalysis (the size of each study is indicated by the size of the solid circle); dark line = weighted regression; and dashed line = unweighted regression. SROC curves summarize the overall diagnostic accuracy.

Sometimes, a differential diagnosis of TPE mandates the use of more invasive procedures like thoracoscopy or thoracotomy. These procedures, which require expertise, may cause complications and may even increase the morbidity of the patients.

Pleural levels of a number of biomarkers have been proposed as aids in the diagnosis of TPE, including those of INF-γ, adenosine deaminase, interleukin-12p40, interleukin-18, immunosuppressive acidic protein, and soluble interleukin-2 receptor, the levels of which are all significantly higher in TPE patients than in non-TPE patients.[45,55] Hiraki and colleagues[49] have compared directly the sensitivities of these markers in a study. ROC analysis demonstrated that INF-γ is the most sensitive and specific indicator of TPE among the above six biological markers (AUC, 1.00). The next most sensitive indicator was soluble interleukin-2 receptor (AUC, 0.99), followed by adenosine deaminase (AUC, 0.96), interleukin-18 (AUC, 0.95), immunosuppressive acidic protein (AUC, 0.93), and interleukin-12p40 (AUC, 0.87). The SROC curve and its AUC present an overall summary of test performance, and display the tradeoff between sensitivity and specificity. The present metaanalysis has shown that the mean sensitivity of the IFN-γ assay was 0.89 while the mean specificity was 0.97, and that the maximum joint sensitivity and specificity (Q value) was 0.95 while the AUC was 0.99, indicating a high level of overall accuracy. We also noted that only one study[19]

**Table 3—*Weighted Metaregression of the Effects of Methodological Quality and Study Design on Diagnostic Precision of Pleural IFN-γ in 22 Assays***

| Covariates | Studies, No. | Coefficient | RDOR (95% CI) | p Value |
|---|---|---|---|---|
| STARD ≥ 13 | 13 | –0.090 | 0.91 (0.16–5.36) | 0.915 |
| QUADAS ≥ 10 | 11 | –0.147 | 0.86 (0.15–4.93) | 0.859 |
| Cross-sectional design | 10 | –0.569 | 0.57 (0.13–2.45) | 0.419 |
| Consecutive or random | 13 | 0.690 | 1.99 (0.33–12.23) | 0.428 |
| Blinded | 12 | 0.757 | 2.13 (0.22–20.89) | 0.488 |
| Prospective | 14 | 0.875 | 2.40 (0.35–16.48) | 0.346 |

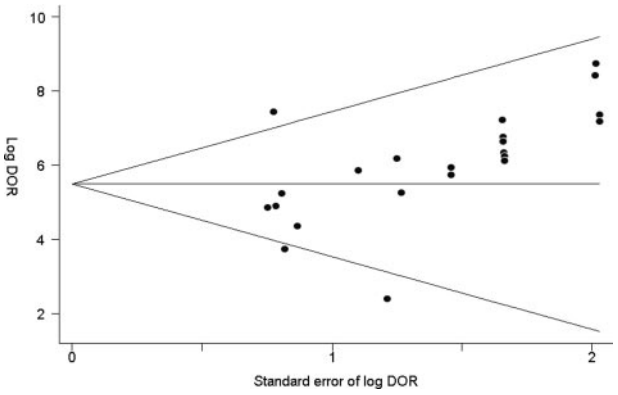


FIGURE 3. Funnel graph for the assessment of potential publication bias in IFN-γ assays. The funnel graph plots the log of the DOR against the SE of the log of the DOR (an indicator of sample size). ● = each study in the metaanalysis; center line = SDOR. The result of the Egger test for publication bias was significant (p = 0.023).

showed relatively low sensitivity (< 0.70) and low specificity (< 0.90) for the detection of IFN-γ in diagnosing TPE, which had low START and QUADAS scores with the smallest study size (n = 21) among all studies included in the present metaanalysis. This situation might account for the low sensitivity and low specificity of this study.

The DOR is a single indicator of test accuracy[56] that combines the data from sensitivity and specificity into a single number. The DOR of a test is the ratio of the odds of positive test results in the patient with disease relative to the odds of positive test results in the patient without disease. The value of a DOR ranges from 0 to infinity, with higher values indicating better discriminatory test performance (*ie*, higher accuracy). A DOR of 1.0 indicates that a test does not discriminate between patients with the disorder and those without it. In the present metaanalysis, we have found that the mean DOR was 272.7, also indicating a high level of overall accuracy.

Since the SROC curve and the DOR are not easy to interpret and use in clinical practice,[57] and since likelihood ratios are considered to be more clinically meaningful,[57,58] we also presented both PLR and NLR as our measures of diagnostic accuracy. Likelihood ratios of > 10 or < 0.1 generate large and often conclusive shifts from pretest to posttest probability (indicating high accuracy).[58] A PLR value of 23.45 suggests that patients with TPE have an approximately 23-fold higher chance of being IFN-γ assay-positive compared with patients without TPE. This high probability would be considered high enough to begin or to continue antituberculosis treatment of TPE patients, especially in the case of the absence of any malignant evidences. On the other hand, NLR was found to be 0.11 in the present metaanalysis. If the IFN-γ assay result was negative, the probability that this patient has TPE is approximately 10%, which is not low enough to rule out TPE. These data suggest that a negative IFN-γ assay result should not be used alone as a justification to deny or to discontinue antituberculosis therapy. The choice of therapeutic strategy should be based on the results of microscopic examination of a smear or culture for *M tuberculosis* and/or histologic observation of pleural tissue, as well as the other clinical data, such as the response to antituberculosis therapy.

An exploration of the reasons for heterogeneity rather than the computation of a single summary measure is an important goal of metaanalysis.[59] In our metaanalysis, both STARD and QUADAS scores were used in the metaregression analysis to assess the effect of study quality on RDOR. We did not observe that the studies with higher quality (*ie*, STARD score of ≥ 13 or QUADAS score of ≥ 10) had better test performances than those with lower quality. Although we found a significant heterogeneity for sensitivity and NLR among the studies analyzed, we noted that differences for studies with or without blinded, cross-sectional, consecutive/random, and prospective design did not reach statistical significance, indicating that the study design did not substantially affect diagnostic accuracy.

It should be emphasized that a definite diagnosis of TPE is achieved when *M tuberculosis* is demonstrated in sputum or pleural specimens, or when caseating granulomas are found in pleural biopsy specimens. As abovementioned, microscopy of the pleural fluid is rarely positive (< 5%).[50–52] Histologic examination of pleura by needle biopsy is not conclusive in 20 to 40% of patients with TPE.[50,51] When the pleural biopsy finding is negative, mycobacteria can be cultured in pleural specimens in < 10% of patients,[50] and this usually takes at least 3 weeks. Where diagnostic difficulty exists, measuring the levels of several biomarkers, such as adenosine deaminase and IFN-γ, in pleural fluid is useful, and clinicians can embark on empirical antituberculosis therapy while awaiting culture results, especially in young patients from areas with a high prevalence of tuberculosis. One criticism of the use biomarkers rather than cultures for the diagnosis of TPE is that culture results are not available to guide antituberculosis therapy. First and foremost among the shortcomings of this is the fact that none of the biomarkers, including IFN-γ, provide culture and sensitivity data. Culture results are particularly useful if drug-resistant tuberculosis is prevalent.[60]

Our data were consistent with the results of the previous metaanalysis on the accuracy of IFN-γ assays by Greco and colleagues.[5] Their metaanalysis

of 13 studies showed that both sensitivity and specificity estimates were heterogeneous. First of all, we updated that previous metaanalysis by including more recent related studies. An important strength of our study was its comprehensive search strategy. Screening, study selection, and quality assessment were done independently and reproducibly by two reviewers. Data extraction and quality assessment were performed in a blinded fashion to reduce bias. We reduced the problem of missing data by contacting authors. We also explored heterogeneity and potential publication bias in accordance with published guidelines.

Our metaanalysis had several limitations. First, the exclusion of conference abstracts, letters to the editors, and non–English-language studies may have led to publication bias, which was found in the present metaanalysis. However, a review of these abstracts and letters suggests that the overall results were similar to the results in the included English-language studies. Publication bias may also be introduced by the inflation of diagnostic accuracy estimates since studies that report positive results are more likely to be accepted for publication. Second, misclassification bias can occur. TPE is not always diagnosed by either histologic or microbiological examination. Actually, TPE was diagnosed in some patients with TPE based just on the clinical course. This issue regarding accuracy of diagnosis can cause nonrandom misclassification, leading to biased results.

In conclusion, current evidence suggests a potential role for IFN-γ assays in confirming a diagnosis of TPE. Since none of pleural fluid biomarkers, including IFN-γ, is specific for TPE, the results of IFN-γ assays should be interpreted in parallel with clinical findings and the results of conventional tests including microbiological examination and pleural biopsy.

ACKNOWLEDGMENT: We are grateful to the following authors who sent additional information on their primary studies: A. Kaya, C. F. Wong, E. Ribera, K. Ogawa, K. Shimokata, L. Valdés, N. G. Saravia, S. K. Sharma, S. Wongtim, T. Morimoto, T. Pettersson, V. Villena, W. H. Hsu, Y. C. Kim, Y. Aoki, Y. Yamada, and Z. C. Gao.

## References

1 National Technical Steering Group of the Epidemiological Sampling Survey for Tuberculosis. Report on fourth national epidemiological sampling survey of tuberculosis [in Chinese]. Chin J Tuberc Respir Dis 2002; 25:3–7
2 Dye C. Global epidemiology of tuberculosis. Lancet 2006; 367:938–940
3 Ellner JJ, Barnes PF, Wallis RS, et al. The immunology of tuberculous pleurisy. Semin Respir Infect 1988; 3:335–342
4 Light RW. Pleural effusion. N Engl J Med 2002; 346:1971–1977
5 Greco S, Girardi E, Masciangelo R, et al. Adenosine deaminase and interferon γ measurements for the diagnosis of tuberculous pleurisy: a meta-analysis. Int J Tuberc Lung Dis 2003; 7:777–786
6 Bossuyt PM, Reitsma JB, Bruns DE, et al. Standards for reporting of diagnostic accuracy: towards complete and accurate reporting of studies of diagnostic accuracy; the STARD initiative. BMJ 2003; 326:41–44
7 Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003; 3:25
8 Deville WL, Buntinx F, Bouter LM, et al. Conducting systematic reviews of diagnostic studies: didactic guidelines. BMC Med Res Methodol 2002; 2:9
9 Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med 1993; 12:1293–1316
10 Lau J, Ioannidis JP, Balk EM, et al. Diagnosing acute cardiac ischemia in the emergency department: a systematic review of the accuracy and clinical effect of current technologies. Ann Emerg Med 2001; 37:453–460
11 Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med 1994; 120:667–676
12 Vamvakas EC. Meta-analyses of studies of the diagnostic accuracy of laboratory tests: a review of the concepts and methods. Arch Pathol Lab Med 1998; 122:675–686
13 Suzuki S, Moro-oka T, Choudhry NK. The conditional relative odds ratio provided less biased results for comparing diagnostic test accuracy in meta-analyses. J Clin Epidemiol 2004; 57:461–469
14 Westwood ME, Whiting PF, Kleijnen J. How does study quality affect the results of a diagnostic meta-analysis? BMC Med Res Methodol 2005; 8:20
15 Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997; 315:629–634
16 Ribera E, Ocana I, Martinez-Vazquez JM, et al. High level of interferon γ in tuberculous pleural effusion. Chest 1988; 93:308–311
17 Hsu WH, Chiang CD, Chen WT, et al. Diagnostic value of adenosine deaminase and γ-interferon in tuberculous and malignant pleural effusions. Taiwan Yi Xue Hui Za Zhi 1989; 88:879–882
18 Shimokata K, Saka H, Murate T, et al. Cytokine content in pleural effusion: comparison between tuberculous and carcinomatous pleurisy. Chest 1991; 99:1103–1107
19 Maeda J, Ueki N, Ohkawa T, et al. Local production and localization of transforming growth factor-β in tuberculous pleurisy. Clin Exp Immunol 1993; 92:32–38
20 Valdes L, San Jose E, Alvarez D, et al. Diagnosis of tuberculous pleurisy using the biologic parameters adenosine deaminase, lysozyme, and interferon γ. Chest 1993; 103:458–465
21 Aoki Y, Katoh O, Nakanishi Y, et al. A comparison study of IFN-γ, ADA, and CA125 as the diagnostic parameters in tuberculous pleuritis. Respir Med 1994; 88:139–143
22 Soderblom T, Nyberg P, Teppo AM, et al. Pleural fluid interferon-γ and tumour necrosis factor-α in tuberculous and rheumatoid pleurisy. Eur Respir J 1996; 9:1652–1655
23 Kim YC, Park KO, Bom HS, et al. Combining ADA, protein and IFN-γ best allows discrimination between tuberculous and malignant pleural effusion. Korean J Intern Med 1997; 12:225–231
24 Ogawa K, Koga H, Hirakata Y, et al. Differential diagnosis of tuberculous pleurisy by measurement of cytokine concentrations in pleural effusion. Tuber Lung Dis 1997; 78:29–34

25 Wongtim S, Silachamroon U, Ruxrungtham K, et al. Interferon γ for diagnosing tuberculous pleural effusions. Thorax 1999; 54:921–924
26 Villegas MV, Labrada LA, Saravia NG. Evaluation of polymerase chain reaction, adenosine deaminase, and interferon-γ in pleural fluid for the differential diagnosis of pleural tuberculosis. Chest 2000; 118:1355–1364
27 Yamada Y, Nakamura A, Hosoda M, et al. Cytokines in pleural liquid for diagnosis of tuberculous pleurisy. Respir Med 2001; 95:577–581
28 Aoe K, Hiraki A, Murakami T, et al. Diagnostic significance of interferon-γ in tuberculous pleural effusions. Chest 2003; 123:740–744
29 Villena V, Lopez-Encuentra A, Pozo F, et al. Interferon γ levels in pleural fluid for the diagnosis of tuberculosis. Am J Med 2003; 115:365–370
30 Wong CF, Yew WW, Leung SK, et al. Assay of pleural fluid interleukin-6, tumor necrosis factor-α and interferon-γ in the diagnosis and outcome correlation of tuberculous effusion. Respir Med 2003; 97:1289–1295
31 Poyraz B, Kaya A, Ciledag A, et al. Diagnostic significance of γ-interferon in tuberculous pleurisy. Tuberk Toraks 2004; 52:211–217
32 Sharma SK, Banga A. Diagnostic utility of pleural fluid IFN-γ in tuberculosis pleural effusion. J Interferon Cytokine Res 2004; 24:213–217
33 El-Ansary AK, Radwan MA. Evaluation of cytokines in pleural fluid for the differential diagnosis of tuberculous pleurisy. Biochim Clin 2005; 29:13–20
34 Gao ZC, Tian RX. Clinical investigation on diagnostic value of interferon-γ, interleukin-12 and adenosine deaminase isoenzyme for tuberculous pleurisy. Chin Med J (Engl) 2005; 118:234–237
35 Okamoto M, Kawabe T, Iwasaki Y, et al. Evaluation of interferon-γ, interferon-γ-inducing cytokines, and interferon-γ-inducible chemokines in tuberculous pleural effusions. J Lab Clin Med 2005; 145:88–93
36 Sharma SK, Banga A. Pleural fluid interferon-γ and adenosine deaminase levels in tuberculosis pleural effusion: a cost-effectiveness analysis. J Clin Lab Anal 2005; 19:40–46
37 Morimoto T, Takanashi S, Hasegawa Y, et al. Level of antibodies against mycobacterial glycolipid in the effusion for diagnosis of tuberculous pleural effusion. Respir Med 2006; 100:1775–1780
38 Barnes PF, Fong SJ, Brennan PJ, et al. Local production of tumor necrosis factor and IFN-γ in tuberculous pleuritis. J Immunol 1990; 145:149–154
39 Barbosa T, Arruda S, Chalhoub M, et al. Correlation between interleukin-10 and in situ necrosis and fibrosis suggests a role for interleukin-10 in the resolution of the granulomatous response of tuberculous pleurisy patients. Microbes Infect 2006; 8:889–897
40 Yanagawa H, Yano S, Haku T, et al. Interleukin-1 receptor antagonist in pleural effusion due to inflammatory and malignant lung disease. Eur Respir J 1996; 9:1211–1216
41 Oshikawa K, Yanagisawa K, Ohno S, et al. Expression of ST2 in helper T lymphocytes of malignant pleural effusions. Am J Respir Crit Care Med 2002; 165:1005–1009
42 Chen YM, Yang WK, Whang-Peng J, et al. An analysis of cytokine status in the serum and effusions of patients with tuberculous and lung cancer. Lung Cancer 2001; 31:25–30
43 Oshikawa K, Sugiyama Y. Elevated soluble CD26 levels in patients with tuberculous pleurisy. Int J Tuberc Lung Dis 2001; 5:868–872
44 Kim YK, Lee SY, Kwon SS, et al. γ-interferon and soluble interleukin 2 receptor in tuberculous pleural effusion. Lung 2001; 179:175–184
45 Okamoto M, Hasegawa Y, Hara T, et al. T-helper type 1/T-helper type 2 balance in malignant pleural effusions compared to tuberculous pleural effusions. Chest 2005; 128: 4030–4035
46 Villena V, Lopez-Encuentra A, Echave-Sustaeta J, et al. Interferon-γ in 388 immunocompromised and immunocompetent patients for diagnosing pleural tuberculosis. Eur Respir J 1996; 9:2635–2639
47 Sharma SK, Mitra DK, Balamurugan A, et al. Cytokine polarization in miliary and pleural tuberculosis. J Clin Immunol 2002; 22:345–352
48 Hiraki A, Aoe K, Matsuo K, et al. Simultaneous measurement of T-helper 1 cytokines in tuberculous pleural effusion. Int J Tuberc Lung Dis 2003; 7:1172–1177
49 Hiraki A, Aoe K, Eda R, et al. Comparison of six biological markers for the diagnosis of tuberculous pleuritis. Chest 2004; 125:987–989
50 Escudero Bueno C, Garcia Clemente M, Cuesta Castro B, et al. Cytologic and bacteriologic analysis of fluid and pleural biopsy specimens with Cope's needle: study of 414 patients. Arch Intern Med 1990; 150:1190–1194
51 Epstein DM, Kline LR, Albelda SM, et al. Tuberculous pleural effusions. Chest 1987; 91:106–109
52 Valdes L, Alvarez D, San Jose E, et al. Tuberculous pleurisy: a study of 254 patients. Arch Intern Med 1998; 158:2017–2021
53 Seibert AF, Haynes J Jr, Middleton R, et al. Tuberculous pleural effusion: twenty-year experience. Chest 1991; 99:883–886
54 Perez-Rodriguez E, Jimenez Castro D. The use of adenosine deaminase and adenosine deaminase isoenzymes in the diagnosis of tuberculous pleuritis. Curr Opin Pulm Med 2000; 6:259–266
55 Wong PC. Management of tuberculous pleuritis: can we do better? Respirology 2005; 10:144–148
56 Glas AS, Lijmer JG, Prins MH, et al. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 2003; 56:1129–1135
57 Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London, UK: BMJ Publishing Group, 2001; 248–282
58 Jaeschke R, Guyatt G, Lijmer J. Diagnostic tests. In: Guyatt G, Rennie D, eds. Users' guides to the medical literature: a manual for evidence-based clinical practice. Chicago, IL: AMA Press, 2002; 121–140
59 Petitti DB. Approaches to heterogeneity in meta-analysis. Stat Med 2001; 20:3625–3633
60 Light RW. Establishing the diagnosis of tuberculous pleuritis. Arch Intern Med 1998; 158:1967–1968