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; 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 (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 (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. 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. The sensitivity and specificity for the single test threshold identified for each study were used to plot an SROC curve. A random-effects model was used to calculate the average sensitivity, specificity, and the other measures across studies. The term heterogeneity when used in relation to metaanalyses refers to the degree of variability in results across studies. We used the χ² and Fisher exact tests to detect statistically significant
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 studies35,37 were excluded because IFN-γ concentration was determined only in TPE patients, two studies40,41 were excluded because they recruited patients with confirmed TPE, four studies42–45 were excluded because they did not allow the calculation of sensitivity or specificity, and four studies46–49 were excluded because they included groups of patients that had been described elsewhere. The data in the study by Villena et al46 were reported in another study by Villena et al29 (V. Villena, MD, personal communication; July 2006). The data in the study by Sharma et al47 were reported in the study by Sharma and Banga32, and those in the studies by Hiraki and colleagues48,49 were reported in the study by Aoe et al.28 Subsequently, 22 studies16–37 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. In two studies,32,36 a small proportion of the patients studied received diagnoses 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 granuloma.

<table>
<thead>
<tr>
<th>Study/Year</th>
<th>Patients, No.</th>
<th>Assay Method</th>
<th>Cutoff</th>
<th>Test Results</th>
<th>Quality Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP</td>
<td>FP</td>
</tr>
<tr>
<td>Ribera et al19/1988</td>
<td>80</td>
<td>RIA</td>
<td>2 IU/mL</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>Hsu et al22/1989</td>
<td>39</td>
<td>ELISA</td>
<td>10 IU/mL</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>Shimokata et al24/1991</td>
<td>40</td>
<td>RIA</td>
<td>Unknown</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>Maeda et al25/1993</td>
<td>21</td>
<td>ELISA</td>
<td>300 pg/mL</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Valdes et al26/1993</td>
<td>145</td>
<td>ELISA</td>
<td>140 pg/mL</td>
<td>33</td>
<td>9</td>
</tr>
<tr>
<td>Aoki et al27/1994</td>
<td>39</td>
<td>ELISA</td>
<td>0.3 IU/mL</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Soderblom et al28/1996</td>
<td>102</td>
<td>ELISA</td>
<td>1.5 pg/mL</td>
<td>43</td>
<td>0</td>
</tr>
<tr>
<td>Kim et al29/1997</td>
<td>70</td>
<td>RIA</td>
<td>9.1 IU/mL</td>
<td>29</td>
<td>2</td>
</tr>
<tr>
<td>Ogawa et al30/1997</td>
<td>50</td>
<td>ELISA</td>
<td>5 IU/mL</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>Wongkin et al31/1999</td>
<td>66</td>
<td>ELISA</td>
<td>240 pg/mL</td>
<td>37</td>
<td>1</td>
</tr>
<tr>
<td>Villegas et al32/2000</td>
<td>137</td>
<td>ELISA</td>
<td>6 IU/mL</td>
<td>45</td>
<td>2</td>
</tr>
<tr>
<td>Yamada et al33/2001</td>
<td>70</td>
<td>RIA</td>
<td>3.1 IU/mL</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>Aoe et al34/2003</td>
<td>46</td>
<td>ELISA</td>
<td>5.7 IU/mL</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Villena et al35/2003</td>
<td>505</td>
<td>RIA</td>
<td>3.7 IU/mL</td>
<td>80</td>
<td>12</td>
</tr>
<tr>
<td>Wong et al36/2003</td>
<td>66</td>
<td>ELISA</td>
<td>60 pg/mL</td>
<td>32</td>
<td>0</td>
</tr>
<tr>
<td>Poyraz et al37/2004</td>
<td>45</td>
<td>ELISA</td>
<td>12 pg/mL</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>Sharma and Banga38/2004</td>
<td>101</td>
<td>ELISA</td>
<td>138 pg/mL</td>
<td>58</td>
<td>1</td>
</tr>
<tr>
<td>El-Assary and Radwan39/2005</td>
<td>39</td>
<td>ELISA</td>
<td>3.1 IU/mL</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>Gao and Tian40/2005</td>
<td>190</td>
<td>ELISA</td>
<td>61.7 pg/mL</td>
<td>119</td>
<td>2</td>
</tr>
<tr>
<td>Okamoto et al41/2005</td>
<td>43</td>
<td>ELISA</td>
<td>99.3 pg/mL</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Sharma and Banga42/2005</td>
<td>52</td>
<td>ELISA</td>
<td>167.5 pg/mL</td>
<td>34</td>
<td>0</td>
</tr>
<tr>
<td>Morimoto et al43/2006</td>
<td>65</td>
<td>ELISA</td>
<td>248 pg/mL</td>
<td>16</td>
<td>3</td>
</tr>
</tbody>
</table>

*ELISA = enzyme-linked immunosorbent assay; RIA = radioimmunoassay; TP = true-positive; FP = false-positive; FN = false-negative; TN = true-negative.
loma. In the remaining 19 studies, 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. 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], 0.57 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). *χ²* 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*

<table>
<thead>
<tr>
<th>Study/Year</th>
<th>TB Patients, No./Non-TB Subjects, No.</th>
<th>Reference Standard</th>
<th>Cross-Sectional Design</th>
<th>Consecutive or Random Design</th>
<th>Blinded Design</th>
<th>Prospective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ribera et al16/1988</td>
<td>30/50 Bac/His</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Hsu et al17/1989</td>
<td>19/20 Bac/His</td>
<td>Unknown</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Shimokata et al18/1991</td>
<td>20/20 Bac/His</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Maeda et al19/1993</td>
<td>14/7 Bac/His</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>Valdes et al20/1993</td>
<td>35/110 Bac/His</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Aoki et al21/1994</td>
<td>11/28 Bac/His</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Soderblom et al22/1996</td>
<td>54/48 Bac/His or Clin</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Kim et al23/1997</td>
<td>39/31 Bac/His</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Ogawa et al24/1997</td>
<td>18/32 Bac/His</td>
<td>Unknown</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Wongtin et al25/1999</td>
<td>39/27 Bac/His</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Villagas et al26/2000</td>
<td>58/79 Bac/His or Clin</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Yamada et al27/2001</td>
<td>21/49 Bac/His</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Aoe et al28/2003</td>
<td>10/36 Bac/His</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Villena et al29/2003</td>
<td>82/513 Bac/His</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Wong et al30/2003</td>
<td>32/34 Bac/His</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Poyraz et al31/2004</td>
<td>15/30 Bac/His</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Sharma and Banga32/2004</td>
<td>64/37 Bac/His</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>El-Ansary and Radwan33/2005</td>
<td>15/24 Bac/His</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Gao and Tian34/2005</td>
<td>141/49 Clin</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Okamoto et al35/2005</td>
<td>11/32 Bac/His</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Sharma and Banga36/2005</td>
<td>35/17 Bac/His</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Morimoto et al37/2006</td>
<td>19/46 Bac/His</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

*TB = tuberculosis; Bac = bacteriology; His = histology; Clin = clinical course.*
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, ≥13; QUADAS score, ≥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
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 colleagues49 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 study19

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

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Studies, No.</th>
<th>Coefficient</th>
<th>RDOR (95% CI)</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>STARD ≥ 13</td>
<td>13</td>
<td>-0.090</td>
<td>0.91 (0.16–5.36)</td>
<td>0.915</td>
</tr>
<tr>
<td>QUADAS ≥ 10</td>
<td>11</td>
<td>-0.147</td>
<td>0.86 (0.15–4.93)</td>
<td>0.859</td>
</tr>
<tr>
<td>Cross-sectional design</td>
<td>10</td>
<td>-0.569</td>
<td>0.57 (0.13–2.45)</td>
<td>0.419</td>
</tr>
<tr>
<td>Consecutive or random</td>
<td>13</td>
<td>0.690</td>
<td>1.99 (0.33–12.23)</td>
<td>0.428</td>
</tr>
<tr>
<td>Blinded</td>
<td>12</td>
<td>0.757</td>
<td>2.13 (0.22–20.89)</td>
<td>0.488</td>
</tr>
<tr>
<td>Prospective</td>
<td>14</td>
<td>0.875</td>
<td>2.40 (0.35–16.48)</td>
<td>0.346</td>
</tr>
</tbody>
</table>
showed relatively low sensitivity (<0.70) and low specificity (<0.90) for the detection of IFN-\(\gamma\) 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 \(ie,\) 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-\(\gamma\) 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-\(\gamma\) 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-\(\gamma\) 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 \(\geq 13\) or QUADAS score of \(\geq 10\) 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-\(\gamma\), 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-\(\gamma\), 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-\(\gamma\) 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, Š. 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
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