Diagnostic value of interferon-gamma in tuberculous pleurisy: a metaanalysis

Jiang J, Shi HZ, Liang QL, Qin SM, Qin XJ

CRD summary
This review assessed the accuracy of interferon-gamma in pleural effusions for the diagnosis of tuberculous pleurisy. The authors concluded that it is a sensitive and specific test. The review suffered from limitations relating to the literature search, quality assessment and synthesis, and very few individual study details were presented. These findings should therefore be interpreted with extreme caution.

Authors' objectives
To determine the accuracy of interferon-gamma (IFN-g) for the diagnosis of tuberculous pleurisy (TPE).

Searching
MEDLINE (1980 to 2006), EMBASE (1980 to 2006), Web of Science (1990 to 2006), BIOSIS Previews (1993 to 2006), LILACS (1980 to 2006) and the Cochrane Library were searched; the search terms, which were listed, included a diagnostic filter. The reference lists of primary studies and review articles were screened and experts in the area were contacted for additional relevant studies. No language restrictions were applied to the searches but only English language studies were included in the review.

Study selection
Study designs of evaluations included in the review
Diagnostic accuracy studies that included at least 10 patients with TPE were eligible for inclusion. The included studies were both prospective and retrospective, and some used a cross-sectional design.

Specific interventions included in the review
Studies that evaluated IFN-g for the diagnosis of TPE were eligible for inclusion. Assay methods for the detection of IFN-g were radioimmunoassay or enzyme-linked immunosorbent assay.

Reference standard test against which the new test was compared
Inclusion criteria were not defined in terms of the reference standard. The reference standards used in the included studies were bacteriology, histology or clinical course.

Participants included in the review
Inclusion criteria were not defined in terms of the participants. Details of the participants in the included studies were not reported.

Outcomes assessed in the review
Studies that reported the sensitivity and specificity, or individual patient test results, were eligible for inclusion. The outcomes reported in the review were the sensitivity, specificity, positive and negative likelihood ratios (LRs), and diagnostic odds ratios (DORs).

How were decisions on the relevance of primary studies made?
Two reviewers independently assessed studies for relevance; any disagreements were resolved through consensus.

Assessment of study quality
Two reviewers independently assessed methodological quality using the STARD guidelines for reporting test accuracy studies and the QUADAS (Quality Assessment of Diagnostic Accuracy Studies) tool. Any disagreements were resolved through consensus. The studies were assigned a score out of 25 and another out of 14, according to the number of STARD and QUADAS items fulfilled. If the studies did not report sufficient information to assess any of the quality...
items, authors were contacted for further information. If the authors did not respond then 'unclear' items were treated as 'no'.

**Data extraction**
Two reviewers independently extracted data on the sensitivity, specificity and threshold for defining a positive IFN-g result. Where individual patient IFN-g results were provided graphically, values were extracted from the graphs using a scalar grid and used to produce a receiver operating characteristic (ROC) plot for each study. Any disagreements were resolved through consensus. The sensitivity, specificity, positive and negative LRs, and DORs were calculated for each study.

**Methods of synthesis**

How were the studies combined?
Pooled sensitivity, specificity, positive and negative LRs, and DORs were calculated using random-effects models. A summary ROC analysis was also conducted. Publication bias was investigated using funnel plots and Egger’s test.

How were differences between studies investigated?
Heterogeneity was assessed statistically using the chi-squared statistic and the Fisher exact test. The effects of STARD and QUADAS scores, and other variables, on estimates of accuracy was investigated by including these in univariate meta-regression analyses based on the DOR.

**Results of the review**

Twenty-two studies (2,101 patients) were included.

The QUADAS scores ranged from 6 to 13 out of 14. Fourteen studies were prospective, ten used a cross-sectional design, and the design of the other studies was unclear. Thirteen studies enrolled consecutive patients and twelve reported blind interpretation of the IFN-g assay.

The sensitivity ranged from 64 to 100% and the specificity from 86 to 100%. All but one study reported a specificity greater than 90%. The pooled sensitivity was 89% (95% confidence interval, CI: 87, 91) and the pooled specificity 97% (95% CI: 96, 98). There was strong evidence of heterogeneity for both measures (p<=0.05).

The summary ROC curve was positioned towards the upper left hand corner of the ROC space, suggesting good accuracy, and the weighted area under the curve was 0.98.

None of the meta-regression analyses carried out showed any significant associations between the variables investigated and the DOR (p>0.3).

There was some evidence of publication bias based on the funnel plot and Egger's test (p=0.023).

**Authors' conclusions**
The measurement of IFN-G in pleural effusions is a sensitive and specific test and is likely to be a useful tool for the diagnosis of TPE.

**CRD commentary**
The review addressed a focused question with selection criteria defined in terms of the index test, target condition and outcomes. Inclusion criteria were not defined in terms of the reference standard, population or study design. The literature search included an appropriate range of databases but included a diagnostic filter, and the review was restricted to studies published in English. It is therefore likely that relevant studies have been missed and the review may be subject to language and publication bias. Details of the review process were reported, and these included appropriate steps to minimise errors and bias. A full quality assessment was carried out using appropriate criteria for test accuracy studies. However, the results were simply presented as summary quality scores, which are problematic,
and the QUADAS tool specifically advises against the use of summary quality scores.

Very few details of the included studies were presented which, combined with the lack of results for the individual quality items, make it very difficult to assess the validity of the included studies and the generalisability of the review findings. The methods used to pool the studies were adequate, although the use of more sophisticated methods would have been preferable. The inclusion of a summary ROC plot greatly helped in the interpretation of the results. Although heterogeneity was investigated, the methods used for this have some limitations and the use of the summary quality scores in such analyses is not appropriate. The reviewers attempted to assess publication bias but the methods used were not appropriate for test accuracy studies. Although the conclusions are supported by the results presented, they should be interpreted with extreme caution given the limitations highlighted.

**Implications of the review for practice and research**
Practice: The authors stated that the IFN-g test is likely to be useful for the diagnosis of TPE, but the results of such tests should be interpreted together with clinical findings and the results of conventional tests.

Research: The authors did not state any implications for further research.

**Funding**
National Natural Science Foundation of China, grant number 30660064; New Century Excellent Talents in Chinese Universities, program number NCET-04-0835; Natural Science Foundation of Guangxi Zhuang Autonomous Zone, China, grant number 0639044.

**Bibliographic details**

**PubMedID**
17426220

**DOI**
10.1378/chest.06-2273

**Original Paper URL**

**Indexing Status**
Subject indexing assigned by NLM

**MeSH**
Biomarkers /metabolism; Confidence Intervals; Diagnosis, Differential; Humans; Interferon-gamma /metabolism; Pleural Effusion /diagnosis /etiology /metabolism; Reproducibility of Results; Sensitivity and Specificity; Tuberculosis, Pleural /complications /diagnosis /metabolism

**AccessionNumber**
12007005489

**Date bibliographic record published**
30/04/2008

**Date abstract record published**
16/05/2008
Record Status
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.