Performance of serum marker panels for liver fibrosis in chronic hepatitis C
Parkes J, Guha I N, Roderick P, Rosenberg W

CRD summary
This review found that serum markers can be used to rule in or rule out fibrosis in up to 35% of patients but cannot be reliably used to stage fibrosis in patients with chronic hepatitis C. The conclusions about the inability of serum markers to stage fibrosis are likely to be reliable, but those relating to the ability to rule fibrosis in or out should be interpreted with caution given the way in which this was calculated.

Authors' objectives
To determine the accuracy of panels of serum markers for the diagnosis of hepatic fibrosis in patients with chronic hepatitis C (CHC).

Searching
MEDLINE, EMBASE and the Cochrane Library were searched from 1985 to October 2004; a diagnostic filter was used, but further details of the search terms were not provided. Relevant websites were screened from 2002 to 2004 for conference proceedings and abstracts. The reference lists from relevant articles were checked and experts on test accuracy reviews were consulted. The review was limited to studies reported in English.

Study selection
Diagnostic accuracy studies of at least 30 patients, systematic reviews or meta-analyses, which assessed panels of at least two serum markers, used liver biopsy as the reference standard and reported data separately for interferon-naive patients with CHC, were eligible for inclusion. Studies that did not produce a composite score based on the serum markers were excluded. Serum markers were defined as any measure that could be derived from a blood sample. The included studies assessed ten different panels of serum markers; full details were given. Different systems were used to stage fibrosis based on the liver biopsy result.

The median age of the included patients was 44.5 years (range: 39 to 47) and the median proportion of men was 64% (range: 45 to 71). Four studies prevented CHC risk factors. The proportion of patients with moderate or severe fibrosis was 43% (range: 17 to 80).

The outcomes reported in the review were the area under the receiver operating curve (AUC), sensitivity, specificity, predictive values, positive and negative likelihood ratios (LR+ and LR-, respectively), and diagnostic odds ratios (DORs).

The included studies used either prospective or retrospective diagnostic cohort designs. Some studies used two cohorts of patients: a training set and a validation set, recruited either from the same or different centres.

Two reviewers assessed studies for inclusion.

Assessment of study quality
The studies were assessed for methodological quality using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool and three additional items added by the authors.

The authors did not state how many reviewers performed the validity assessment.

Data extraction
Where possible, the data were extracted as 2×2 tables of test performance and used to calculate the sensitivity, specificity, predictive values, LR+, LR- and DOR at each threshold value. Thresholds that produced clinically useful predictive values with acceptable false negative and positive rates were used. The accuracy of the test was calculated for different fibrosis stages: early versus moderate or severe fibrosis (F0/F1 versus F2/F3/F4) and for cirrhosis versus
no cirrhosis (F0/F1/F2/F3 versus F4).

One reviewer extracted the data, which were checked by a second reviewer. Any disagreements were resolved through consensus or by referral to a third reviewer.

**Methods of synthesis**

Summary receiver operating characteristic (ROC) curves were estimated based on the Moses-Littenberg model for all tests and all thresholds. The summary sensitivity, specificity and DOR were calculated for the summary ROC at the mean of S (S represents the variation of thresholds across studies). The mean AUC was reported, but it is unclear how this was calculated. Heterogeneity was not formally assessed, although the study results were plotted in ROC space. A subgroup analysis based on method of recruitment, quality of biopsy and scoring system was undertaken.

**Results of the review**

Fourteen primary studies (4,321 patients) were included in the review. Ten provided data to calculate accuracy measures at specific thresholds, while others reported ROC curves. Eleven studies presented data validated in a different group of patients than the training set. Two reviews were also identified but all studies included in these were included individually in the review.

All studies fulfilled QUADAS items on representative sample of patients, description of selection criteria, use of an appropriate reference standard, avoidance of partial and differential verification bias, avoidance of incorporation bias, blind interpretation of the index test and reference standard, and reporting of components of the index test. In all studies it was unclear if the reference standard was reproducible, all studies did not report all test results, 9 studies explained reasons for withdrawals, and 6 studies provided details of the formula to calculate the index test score.

Fourteen studies reported on early versus moderate or severe fibrosis. The mean AUC was 0.77 in validation populations and 0.81 in training populations. The LR- ranged from 0.1 to 0.9 and the LR+ from 1.2 to 33.1. At each threshold tests either had high sensitivity and low specificity or vice versa, none showed both high sensitivity and specificity.

Six studies reported on cirrhosis versus no cirrhosis. All studies reported greater accuracy (higher AUC and/or sensitivity and specificity) than for distinguishing early from moderate or severe fibrosis.

**Authors’ conclusions**

Serum markers can be used to rule in or rule out fibrosis in up to 35% of patients but cannot be reliably used to stage fibrosis.

**CRD commentary**

The review was reported clearly and addressed a focused question with defined inclusion criteria. The literature search was limited to two databases and included a diagnostic filter, which might have resulted in relevant studies being missed. Some attempts were made to locate unpublished studies in the form of abstracts and conference proceedings, but the review was limited to English language articles; there is therefore a possibility of both language and publication bias. Appropriate steps were taken to minimise bias and error in the study selection and data extraction of data processes, but it is unclear whether such steps were also taken for the validity assessment. Appropriate criteria were used to assess study quality and the results of this assessment were presented clearly. The studies were generally found to be of similar, good quality and so there was no need to further investigate individual quality items in the analysis. The analysis appeared appropriate but the use of advanced statistical methods, such as the hierarchical summary ROC model, would have been likely to produce more reliable results; limited details were provided for some stages of the analysis such as the methods used to estimate mean AUC and the analysis based on predictive values. For the latter analysis, it appears that predictive values were based on those reported in the individual studies which are highly dependent on the pre-test probability of disease, for which data were not reported for this analysis. The authors’ conclusions about the inability of serum markers to stage fibrosis are likely to be reliable, but those relating to the ability to rule fibrosis in or out should be interpreted with caution given the way in which this was calculated.
Implications of the review for practice and research

Practice: The authors stated that serum markers can be used to rule in or rule out fibrosis in up to 35% of patients but should not be used to stage fibrosis.

Research: The authors stated that improvements in index and reference standard tests are needed, including the evaluation of clinical outcomes as a reference standard. New markers based on methods such as proteomics and metabonomics, and their use in combination with existing or new panels of markers, should be investigated. Improved reporting is needed so that all studies provide data as 2×2 tables of test performance.

Funding
Not stated.

Bibliographic details

PubMedID
16427156

DOI
10.1016/j.jhep.2005.10.019

Indexing Status
Subject indexing assigned by NLM

MeSH
Biomarkers /blood; Disease Progression; Hepatitis C, Chronic /blood /complications; Humans; Liver Cirrhosis /blood /etiology; Prognosis; Severity of Illness Index

AccessionNumber
12006000908

Date bibliographic record published
03/08/2007

Date abstract record published
03/11/2008

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.