Performance of transient elastography for the staging of liver fibrosis: a meta-analysis

Friedrich-Rust M, Ong MF, Martens S, Sarrazin C, Bojunga J, Zeuzem S, Herrmann E

CRD summary
This review concluded that transient elastography had excellent diagnostic accuracy for cirrhosis. This was independent of underlying liver disease, but accuracy for diagnosis of significant fibrosis depended on underlying disease. Because the review considered overall accuracy and not relative numbers of false positives and false negatives, and the analysis was not fully described, these conclusions should be interpreted cautiously.

Authors' objectives
To assess the performance of transient elastography for diagnosis of liver fibrosis.

Searching
MEDLINE, EMBASE, The Cochrane Library and Web of Science were searched from 2002 to April 2007. Search terms were reported. Relevant websites and conference abstracts (American Association for the Study of the Liver, European Association for the Study of the Liver, Digestive Disease Week, Liver Transplantation, Asian Pacific Association for the Study of the Liver, Conference on Retroviral and Opportunistic Infections, Interscience Conference on Antimicrobial Agents and Chemotherapy, and International Symposium on Ultrasonic Imaging and Tissue Characterization) were searched from 2002 to April 2007; presenting authors were contacted for any missing data. Bibliographies of included studies were screened for additional articles.

Study selection
Studies that evaluated transient elastography and used liver biopsy as a reference standard were eligible for inclusion. Included studies were required to use comparable liver biopsy staging systems (METAVIR, Ishak, Brunt, Ludwig’s, Knodell, Desmet or Scheuer), to assess diagnostic accuracy for fibrosis stage (F≥2, F≥3, F=4) according to METAVIR or a comparable staging system and/or assess diagnostic accuracy for a fibrosis stage based on a defined cut-off point for liver stiffness. Diagnostic accuracy data could be reported as sensitivity and specificity or area under the receiver operating characteristic curve (AUROC).

Where reported, mean age of study participants ranged from 11 to 60 years and the proportion of male participants ranged from 8% to 83%. Most of the included studies (36/50) used the METAVIR staging system for liver biopsies. Histopathological diagnoses varied and included: HIV; alcoholic steatohepatitis; non alcoholic steatohepatitis; primary biliary cirrhosis; primary sclerosing cholangitis; non alcoholic fatty liver disease; hepatitis C virus (HCV); hepatitis B virus (HBV); haemochromatosis; and cystic fibrosis.

The authors did not state how many reviewers assessed studies for inclusion.

Assessment of study quality
Methodological quality of included studies was assessed using the QUADAS tool for assessment of aspects reporting quality, appropriateness of participant spectrum, use of an appropriate reference standard, blinding of test interpreters, avoidance of verification biases, avoidance of disease progression bias and accounting for withdrawals and uninterpretable test results; results were published online (supplementary table 1).

The authors did not state how many reviewers performed quality assessment.

Data extraction
Data were extracted on sensitivity and specificity and AUROC of transient elastography for different fibrosis stages (F≥2, F≥3, F=4), based on the METAVIR staging system.

Data extraction was undertaken by one reviewer and checked by a second; disagreements were resolved by discussion.
Methods of synthesis
Studies that used histology scoring systems with a range from 0 to 4 (METAVIR, Desmet and Scheuer, Knodell, Brunt, Ludwig’s) were used to generate a pooled estimate of AUROC and 95% confidence intervals (CIs), using a DerSimonian and Laird random effects model weighted by sample size. Ishak score (scale 0 to 6), was transferred into METAVIR with Ishak F≥3 equivalent to METAVIR F≥2, Ishak F≥4 equivalent to METAVIR F≥3 and Ishak F≥5 equivalent to METAVIR F=4.

Potential causes of between-study heterogeneity were explored: underlying liver disease (studies of patients with hepatitis C virus only, studies with mixed populations that included hepatitis C virus, studies that excluded hepatitis C virus); staging system used; country; publication status; mean body mass index of participants; mean age of participants; fibrosis stage; gender distribution; mean or median length of liver biopsy specimen; proportion of liver biopsy failures; proportion of FibroScan (transient elastography) failures; and QUADAS criteria.

A summary receiver operating characteristic (SROC) curve was calculated from all studies in which sensitivity and specificity were known for at least one cut-off level, using the Littenberg and Moses model weighted by sample size.

Results of the review
Fifty studies (15 full publications and 35 abstracts, participant numbers ranged from 30 to 1,345) were included in the review. Selection criteria, details of the reference standard and blinding of interpreters of the index test and reference standard were poorly reported by most included studies.

Diagnosis of significant fibrosis (F≥2): Pooled AUROC was 0.84 (95% CI 0.82 to 0.86; 35 studies). Underlying liver disease, histological staging system used and the country in which studies were conducted were significant contributors to between-study heterogeneity; other factors were not significant contributors to heterogeneity.

Diagnosis of severe fibrosis (F≥3): Pooled AUROC was 0.89 (95% CI 0.88 to 0.91; 35 studies). The histological staging system used was a significant contributor to between-study heterogeneity and mean participant body mass index was of borderline significance; other factors were not significant contributors to heterogeneity.

Diagnosis of cirrhosis (F=4): Pooled AUROC was 0.94 (95% CI 0.93 to 0.95; 38 studies). The country in which studies were conducted was a significant contributor to between-study heterogeneity and the histological staging system used was of borderline significance; other factors were not significant contributors to heterogeneity.

Analysis of the impact of study quality found that differences in the QUADAS criteria on selection criteria, appropriate reference standard, partial verification bias, reference execution details, test review bias, diagnostic review bias and a number of uninterpretable results were significant contributors to between-study heterogeneity. The sum of all QUADAS items had no significant influence on the AUROC.

Authors’ conclusions
Transient elastography had excellent diagnostic accuracy for cirrhosis, which was independent of the underlying liver disease. However, for diagnosis of significant fibrosis, a high variation of AUROC was found that was dependent on the underlying liver disease.

CRD commentary
The review addressed a clearly stated research question, which was defined by appropriate inclusion criteria. A range of sources, including sources of unpublished data, were searched for relevant studies. No search restrictions were reported. Measures to avoid error and/or bias were taken during the data extraction process; it was unclear whether similar measures were applied throughout the review process. Methodological quality of included studies was assessed and its impact upon the results of the review considered. Potential sources of between-study heterogeneity (in addition to aspects of methodological quality) were considered, but the methods used to determine their significance were not fully reported. Although sensitivity and specificity data appeared to have been available for many of the included studies, analysis focused on pooled estimates of AUROC. Use of this measure of overall accuracy results in a loss of clinically important information about test performance; it was unclear how many inaccurate test results were due to false
positive and how many to false negatives. Generation of SROC curves using a bivariate of hierarchical model may have been more appropriate to this data set; such models allow generation of summary estimates of sensitivity and specificity as well as potential to assess the significance of sources of heterogeneity. The limitations in the analysis described mean that the authors conclusions should be interpreted with caution.

### Implications of the review for practice and research

**Practice:** The authors made no recommendations for practice.

**Research:** The authors stated that large well-conducted randomised trials with clearly defined end points (such as five-year survival without hepatitis C virus-related cirrhosis or complications related to liver disease) were needed to compare transient elastography with liver biopsy and biochemical markers.

### Funding

This study was supported by the Federal Ministry of Education and Research (BMBF) program Kompetenznetz Hepatitis (Hep-Net).

### Bibliographic details


**PubMedID**

18395077

**DOI**

10.1053/j.gastro.2008.01.034

**Original Paper URL**

http://www.gastrojournal.org/article/S0016-5085(08)00108-X/abstract

**Indexing Status**

Subject indexing assigned by NLM

**MeSH**

Diagnosis, Differential; Elasticity Imaging Techniques /methods; Humans; Liver Cirrhosis /diagnosis /physiopathology; ROC Curve; Reproducibility of Results; Severity of Illness Index

**AccessionNumber**

12008103062

**Date bibliographic record published**

01/12/2008

**Date abstract record published**

05/05/2010

**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.