Non-invasive diagnostic assessment tools for the detection of liver fibrosis in patients with suspected alcohol-related liver disease: a systematic review and economic evaluation

M Stevenson,[1]* M Lloyd-Jones,[1] MY Morgan[2] and R Wong[1]

[1]The University of Sheffield, School of Health and Related Research (ScHARR), Sheffield, UK
[2]The Centre for Hepatology, University College London Medical School, London, UK

*Corresponding author

Executive summary

Executive summary

Background

Excessive alcohol consumption may lead to the development of alcohol-related liver disease (ALD). ALD comprises a spectrum of disease, including hepatic steatosis (alcoholic fatty liver), alcoholic hepatitis, alcoholic fibrosis and cirrhosis, and hepatocellular cancer. In 2008, 0.95% of all deaths registered in people aged ≥ 20 years in England and Wales were attributed to ALD. Liver biopsy may be used in patients with suspected ALD to confirm the diagnosis, exclude other or additional liver pathologies, and provide accurate staging of the degree of liver injury in order to enable the prediction of prognosis and inform treatment decisions. However, as it is an invasive procedure that carries the risk of morbidity and mortality, current UK guidance recommends that biopsy is not required to confirm the diagnosis in patients with a high clinical suspicion of ALD in whom blood tests have excluded other causes of liver disease, unless it is necessary to confirm a diagnosis of acute alcoholic hepatitis in order to inform specific treatment decisions.

Objectives

The objectives of this assessment are to evaluate the diagnostic accuracy, cost-effectiveness, and effect on patient outcomes of four non-invasive tests for liver fibrosis [the Enhanced Liver Fibrosis (ELF™) test (Siemens Healthcare Diagnostic Inc., Tarrytown, NY, USA), FibroTest (BioPredictive, Paris, France), FibroMAX (BioPredictive, Paris, France) and transient elastography (FibroScan®; produced by EchoSens, Paris, France and distributed in the UK by Artemis Medical Ltd, Kent, UK)] in patients suspected of having liver fibrosis related to alcohol consumption. The tests are assessed first as a replacement for liver biopsy, and secondly as an additional test prior to liver biopsy.

Methods

A systematic review was undertaken to identify studies reporting the diagnostic and prognostic accuracy of the ELF test, FibroTest, FibroMAX and FibroScan for the identification of liver fibrosis and associated conditions in patients with suspected ALD. The following databases were searched in January 2010: MEDLINE (from 1950 to January 2010), MEDLINE In-Process & Other Non-Indexed Citations (from 1950 to January 2010), EMBASE (from 1980 to January 2010), Cochrane Database of Systematic Reviews (from 1996 to January 2010), Cochrane Central Register of Controlled Trials (from 1898 to January 2010), Cochrane Methodology Register (from 1904 to January 2010), Database of Abstracts of Reviews of Effects (from 1995 to January 2010), HTA Database (from 1995 to January 2010), NHS Economic Evaluation Database (from 1995 to January 2010), Cumulative Index to Nursing and Allied Health Literature (from 1982 to January 2010), Web of Knowledge, Science Citation Index, Conference Proceedings Citation Index, and BIOSIS Previews (from 1969 to January 2010). Research registers and conference proceedings were also searched. Study quality was assessed using the QUADAS (QUality Assessment of Diagnostic Accuracy Studies) checklist.

A mathematical model was constructed to estimate the incremental costs and incremental quality-adjusted life-years (QALYs) associated with the introduction of alternative strategies compared with a biopsy-all strategy. Owing to the wide uncertainty in the data to populate key variables, 36 scenarios were assessed that varied the sensitivity of biopsy, the anxiety associated with biopsy, different values for the sensitivity and specificity for each non-invasive tests, and whether a percutaneous or transjugular biopsy was required. For each of these scenarios, nine strategies were evaluated, which were divided into triage strategies (where a positive test was confirmed by biopsy) and replacement strategies (where no confirmatory biopsy was provided). For each scenario and strategy, two threshold levels were reported where biopsying all patients was more cost-effective than the strategy: the decreased level of abstinence associated with the strategy compared with biopsying all and the level of QALY gain that would be required for a biopsy.

Results

Summary of clinical results

Diagnostic accuracy of the Enhanced Liver Fibrosis Test

No studies were identified that specifically assessed the ELF test. One study evaluated the diagnostic accuracy of the European Liver Fibrosis Test (essentially, the ELF test with the addition of age to the algorithm) compared with liver biopsy in patients with chronic liver disease, only 64 of whom had ALD; a follow-up study in 85 patients with ALD assessed its ability to predict long-term survival and relevant clinical events. This limited evidence suggests that, using a threshold score of 0.431, the European Liver Fibrosis Test can differentiate between moderate/severe fibrosis and milder/no fibrosis in patients with ALD with a sensitivity of 93% and a specificity of 100%; it is less good at identifying cirrhosis. It appears to have some predictive value in relation to both liver-related clinical outcomes and all-cause mortality. However, because the results rest on data from so few patients, evidence for the diagnostic and prognostic accuracy of the test is not robust.

Diagnostic accuracy of FibroTest

Five studies of FibroTest were identified. Two evaluated diagnostic test accuracy compared with liver biopsy in patients with known or suspected ALD. A further three recruited patients with liver disease of mixed aetiology, including ALD: the first assessed FibroTest's ability to identify portal hypertension (PHt) and also compared it with liver biopsy, the second assessed its ability to predict the presence of oesophageal varices, and the third assessed its predictive value in relation to survival at 2 and 6 months in patients with severe cirrhosis. Results from the largest study, which was also the most representative of the spectrum of patients with suspected ALD, suggest that, in such patients, using a threshold score of 0.30, FibroTest can differentiate between moderate/severe fibrosis and milder/no fibrosis with a sensitivity of 84% and specificity of 66%, while using a threshold score of 0.70, it can distinguish cirrhosis with a sensitivity of 91% and specificity of 87%. Very small studies suggest that, using a threshold score of 0.58, FibroTest can distinguish between patients with and without clinically significant PHt with a sensitivity of 93% and specificity of 87%, while, using a threshold score of 0.85, it can distinguish between those with and without grade 2 oesophageal varices with a sensitivity of 89% and specificity of 50%. However, the results relating to PHt and oesophageal varices are not robust because the studies were very small, and the conditions of interest were over-represented. FibroTest appears to predict survival with relatively low accuracy.

Diagnostic accuracy of FibroMAX

No relevant studies of FibroMAX were identified.

Diagnostic accuracy of FibroScan

Six studies were identified that assessed the diagnostic test accuracy of FibroScan relative to liver biopsy in patients with known or suspected ALD. A further three studies recruited patients with liver disease of mixed aetiology, including ALD. One assessed the ability of FibroScan to predict the presence of large oesophageal varices in patients with cirrhosis, whereas the other two assessed its ability to predict clinically significant PHt. The study with the most representative population suggests that, using threshold scores of 5.9, 7.8, 11.0 and 19.5 kPa, respectively, FibroScan can differentiate between patients with and without fibrosis with a sensitivity of 83% and a specificity of 86%, and can identify moderate/severe fibrosis with a sensitivity of 80% and a specificity of 90.5%, severe fibrosis with a sensitivity of 87% and specificity of 80.5%, and cirrhosis with a sensitivity of 86% and specificity of 84%. However, again, these results are not robust because the study was relatively small and the conditions of interest were over-represented. FibroScan appears to be able to distinguish between patients with and without PHt, and with less success between patients with and without large oesophageal varices. There are no long-term data relating FibroScan results to survival or other clinical outcomes.

Adverse effects and contraindications

The non-invasive tests included in this review appear to be safe. The adverse events associated with the ELF test, FibroTest, and FibroMAX are those associated with diagnostic venepuncture generally: primarily pain and bruising, with occasional vasovagal reactions and very rarely potentially disabling nerve injuries. There is no evidence to indicate that FibroScan is specifically associated with any adverse effects. By contrast, liver biopsy is associated with a high level of morbidity and occasional mortality.

No contraindications have been specified for the ELF test. The contraindications specified for FibroTest, FibroMAX, and FibroScan all relate to the mode of operation of the test, and do not relate to any potential for harm in patients with the relevant characteristics, although they will restrict their practical utility. The most important of these limitations is the restriction on the use of FibroScan in obese patients.

Summary of cost-effectiveness and benefits versus risks

It was concluded that no robust estimate could be provided regarding the incremental costs, incremental QALYs and, therefore, the cost per QALY of a strategy. Scenarios exist in which each of the strategies analysed is more cost-effective than biopsying all patients and, in contrast, scenarios exist in which each strategy is less cost-effective than biopsying all patients. No conclusive result can be provided on the most cost-effective strategy until further data are available; however, there is evidence that some strategies, such as using clinical experience or diagnosing all patients with cirrhosis, will not be the most cost-effective.

Conclusions

Implications for service provision

Owing to the lack of a conclusion regarding the cost-effectiveness of the strategies, it is anticipated that there would be no change in service provision.

Suggested research priorities

A large number of parameters require data; however, the following are selected as being of most importance:

- the sensitivity and specificity of liver biopsy against a gold-standard of post-mortem evaluation of fibrosis
- the sensitivity and specificity of each non-invasive liver test (NILT) against a gold standard of post-mortem evaluation of fibrosis (or failing this biopsy at validated and pre-selected cut-off thresholds for the various degrees of liver damage)
- the influence of potential confounding variables such as current drinking behaviour and the degree of hepatic inflammation on the performance of NILTs
- differential information on the percentage of alcohol misusers who will develop alcohol-related cirrhosis over time, by age at onset, gender and ethnic origin
- the likelihood, and magnitude, of decreases in abstinence rates associated with a diagnosis of significant ALD by diagnostic modality
- the incidental gains in QALYs that may be associated with biopsy, because of the determination of non-ALD-related aetiologies.

Funding

Publication

NIHR Health Technology Assessment programme

The Health Technology Assessment (HTA) programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The research findings from the HTA programme directly influence decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC). HTA findings also help to improve the quality of clinical practice in the NHS indirectly in that they form a key component of the 'National Knowledge Service'.

The HTA programme is needs led in that it fills gaps in the evidence needed by the NHS. There are three routes to the start of projects.

First is the commissioned route. Suggestions for research are actively sought from people working in the NHS, from the public and consumer groups and from professional bodies such as royal colleges and NHS trusts. These suggestions are carefully prioritised by panels of independent experts (including NHS service users). The HTA programme then commissions the research by competitive tender.

Second, the HTA programme provides grants for clinical trials for researchers who identify research questions. These are assessed for importance to patients and the NHS, and scientific rigour.

Third, through its Technology Assessment Report (TAR) call-off contract, the HTA programme commissions bespoke reports, principally for NICE, but also for other policy-makers. TARs bring together evidence on the value of specific technologies.

Some HTA research projects, including TARs, may take only months, others need several years. They can cost from as little as £40,000 to over £1 million, and may involve synthesising existing evidence, undertaking a trial, or other research collecting new data to answer a research problem.

The final reports from HTA projects are peer reviewed by a number of independent expert referees before publication in the widely read journal series *Health Technology Assessment*.

Criteria for inclusion in the HTA journal series
Reports are published in the HTA journal series if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.