Journal of Gastroenterology and Hepatology

JGHF

doi:10.1111/jgh.15482

META ANALYSIS AND SYSTEMATIC REVIEW

Systematic review: Accuracy of the enhanced liver fibrosis test for diagnosing advanced liver fibrosis and cirrhosis

Chetanya Sharma,* Sara Cococcia,*,† Nicola Ellis,* Julie Parkes‡ and William Rosenberg*

*Institute for Liver and Digestive Health, University College London, Division of Medicine and Royal Free London NHS Foundation Trust, London, ‡Department of Public Health and Medical Statistics, Faculty of Medicine, University of Southampton, Southampton, UK; †First Department of Internal Medicine, San Matteo Hospital Foundation, University of Pavia, Pavia, Italy

Key words
Diagnostic accuracy, enhanced liver fibrosis test, liver biopsy, liver fibrosis.

Accepted for publication 22 February 2021.

Correspondence
William Rosenberg, Institute for Liver and Digestive Health, Division of Medicine, University College London, Royal Free Campus, Rowland Hill Street, Hampstead, London, NW3 2PF, UK.
Email: w.rosenberg@ucl.ac.uk

Declaration of conflict of interest: WMR has received sponsorship from Gilead Sciences to attend meetings and has served on advisory boards for Gilead Sciences. WMR is an inventor of the ELF test and has received speaker's fees from Siemens Healthineers.
Author contribution: C. S. performed the research. C. S., S. C., and N. E. collected and analyzed the data. C. S., S. C., and N. E. designed the research study and wrote the paper. W. M. R. and J. P. supervised C. S., S. C., and N. E., reviewed the paper, and made final critical revision for important intellectual contents. All authors have approved the final version of this manuscript.
Financial support: W. M. R. is supported by the UCLH NIHR BRC and is a NIHR senior investigator.
Guarantor of the article: W. M. R. is the guarantor for this article.

Abstract

Background and Aims: The rising incidence of chronic liver disease (CLD) has increased the need for early recognition. This systematic review assesses the diagnostic accuracy of the enhanced liver fibrosis (ELF) test in cases of advanced fibrosis and cirrhosis due to multiple etiologies in at-risk populations.
Methods: Studies evaluating the ELF accuracy in identifying advanced fibrosis or cirrhosis, defined as METAVIR stage $F \geq 3$ and $F = 4$ or equivalent, in patients with non-alcoholic fatty liver disease (NAFLD), alcohol liver disease (ALD), or viral hepatitis were included. Liver biopsy was used as the reference standard. Medline and Embase databases were searched. The QUADAS-2 tool was used as a framework to assess risk of bias and applicability. The area under the receiver operator curve (AUROC) was extracted as a summary measure of diagnostic accuracy.
Results: Thirty-six studies were included: 11 hepatitis C, 4 hepatitis B, 9 NAFLD, 2 ALD, and 10 mixed. The ELF test showed good diagnostic performance in detecting advanced fibrosis in patients with viral hepatitis (AUROC 0.69 to 0.98) and excellent performance in NAFLD (AUROC 0.78 to 0.97) and ALD (AUROC from 0.92 to 0.94). There is also evidence of good diagnostic performance for detecting cirrhosis in patients with viral hepatitis (AUROC 0.63 to 0.99), good performance in NAFLD (AUROC 0.85 to 0.92), and excellent performance in patients with ALD (AUROC 0.93 to 0.94).
Conclusion: This systematic review supports the use of the ELF test across a range of CLD as a possible alternative to liver biopsy in selected cases.

Background

Target condition. Chronic liver disease (CLD) is a leading cause of death globally, with liver-related deaths increasing in England compared with other major killers.[1] The commonest causes of CLD are alcohol, obesity, and viral hepatitis. CLD can lead to liver fibrosis characterized by increased synthesis and altered deposition of extracellular matrix. Fibrosis is usually silent until cirrhosis leads to complications of portal hypertension including variceal bleeding, ascites, and hepatocellular carcinoma. Many patients with CLD present when it is too late to prevent these complications, and they can only be ameliorated. There is a need for tests to detect the presence of fibrosis before it causes irreversible damage, to stratify which patients might benefit from specialist care, and to target surveillance for complications.[2]

Liver biopsy is the reference test for assessing liver fibrosis, but its accuracy is limited by sampling error and inter-observer and intra-observer variation.[3,4] Additionally, it is invasive and can

cause harm, making it unsuitable for monitoring changes in fibrosis. Non-invasive tests (NITs) for liver fibrosis have been developed including transient elastography (TE) and serum biomarkers. TE can be performed at the point-of-care, is painless, and does not require sedation. Test performance can be affected by feeding,[5] inflammation, age, and obesity.[6] The newer XL probe improved performance in obese patients, but unreliable results are still observed in 25% of these patients.[7]

Serum biomarker tests can be more reproducible than TE, and assay performance can be standardized in a laboratory setting. They avoid sampling errors and remove the influence of operator performance. The enhanced liver fibrosis (ELF) test combines measurement of three serum biomarkers involved in matrix biology comprising hyaluronic acid (HA), Procollagen III amino terminal peptide (PIIINP) and tissue inhibitor of metalloproteinase 1 (TIMP-1).[8] Measurements of these analytes are combined in an algorithm to generate a unitless score that has been validated as a measure of liver fibrosis and to be prognostic for complications of liver disease.[9] Pre-analytical[10] and analytical assay performance[11] combine to deliver excellent test performance making this one of the best performing serum biomarkers.

Compared with liver biopsy, the ELF test provides a continuous score rather than a categorical variable and so is more sensitive to changes. Unlike TE, the same test thresholds can be applied to staging different CLD etiologies. While some studies suggest that TE can overestimate fibrosis in the presence of steatosis,[12–14] ELF performs well in the presence of steatohepatitis.[15,16] The impact of extrahepatic fibrosis has yet to be fully quantified.[17]

Related literature. Systematic reviews have investigated the diagnostic accuracy of a range of tests for diagnosing liver fibrosis.[18,19] The use of the ELF test and other NITs is recommended in the European Association for the Study of the Liver guidelines for risk stratification of patients with CLD,[20,21] the British Society for Gastroenterology[22] guidance on investigation of abnormal liver function, and NICE guidance on non-alcoholic fatty liver disease (NAFLD).[23]

A meta-analysis of the diagnostic accuracy of the ELF test in a range of CLD etiologies by Xie *et al*.[24] found evidence of good performance with considerable diagnostic value in predicting histological fibrosis stage. The summary areas under the receiver operator curve (AUROC) for detecting severe fibrosis and cirrhosis with ELF were 0.8696 and 0.8770, respectively. However, heterogeneity between the studies makes it difficult to make recommendations. This meta-analysis was published in 2014 and included only nine studies. There has since been a significant body of research published on the diagnostic accuracy of the ELF test for differing degrees of fibrosis in a variety of settings.

More recently, Vali *et al*.[25] conducted a systematic review of the use of ELF in NAFLD and presented evidence of good diagnostic performance in the detection of advanced fibrosis and cirrhosis in cohorts of patients with NAFLD. However, modeling suggested that ELF, in common with other NITs, would not perform well in detecting fibrosis in low prevalence settings.

This systematic review aims to determine the accuracy of the ELF test for diagnosing advanced liver fibrosis and cirrhosis in a variety of CLD etiologies, with liver biopsy as the reference standard.

Methods

The methods and approach to the systematic review followed the Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy guidance on conducting systematic reviews of diagnostic test accuracy.[26]

Eligibility criteria. The inclusion criteria were as follows: primary research cross-sectional studies of diagnostic accuracy that had assessed liver fibrosis in adult participants with CLD caused by NAFLD, alcohol liver disease (ALD) or hepatitis B virus (HBV) or hepatitis C virus (HCV); mixed etiology studies including patients from at least one of the above-mentioned disease etiologies; studies assessing the diagnostic accuracy of the ELF test; single or two-gate/case–control designs; studies published in the English language. Existing systematic reviews, conference abstracts, and pediatric patients (<18 years) were excluded.

The ELF test was the index test, while liver biopsy, regardless of the staging classification, was the reference standard. Studies where TE was the reference standard were excluded. Liver biopsy size was used to determine study quality but was not used to exclude studies.

Search strategy. Searches were performed in MEDLINE using Ovid and EMBASE with the following search terms: Enhanced liver fibrosis OR (ELF AND (liver fibrosis OR cirrhosis OR hepatic fibrosis OR liver adj3 fibrosis OR fibrosis adj3 liver OR liver disease OR NAFLD OR NASH OR hepatitis OR chronic liver OR alcoholic liver)) NOT (epithelial lining fluid OR extremely-low frequency OR extremely low frequency or extremely low-frequency OR elf-2 or elf2 or elf-97 or elf97 or elf-4b or elf4b). The paper proposing the ELF algorithm for the first time was included despite appearing under a different name.

Electronic searches were supplemented by reviewing the reference lists of retrieved articles. The last search was conducted on August 20, 2020.

Data extraction. Search results were exported, and duplicates were removed automatically using EndNote. Three people (C. S., N. E., and S. C.) independently reviewed each title and abstract, identifying relevant studies. Where additional duplicates were identified, the oldest version was kept as the original. Where there was a discrepancy, these were reviewed by all researchers to reach a consensus.

Three researchers (C. S., N. E., and S. C.) agreed the information to be extracted (Table S1). The full text of the remaining studies was retrieved and reviewed by at least two researchers (C. S., S. C., and N. E.). Data were extracted for papers meeting the inclusion criteria independently by the three researchers and then discussed until consensus was reached.

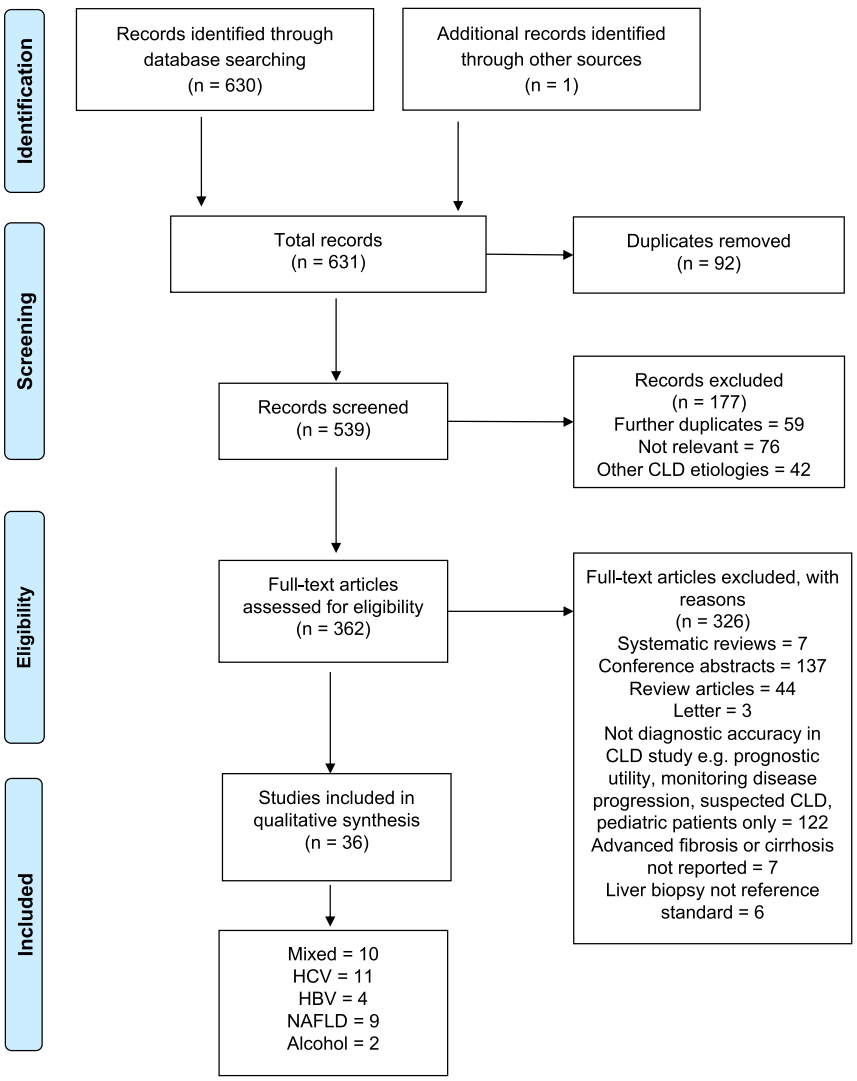
The different histological staging systems used to assess liver fibrosis were aligned to create four categories; any fibrosis; at least moderate fibrosis; at least advanced fibrosis; and cirrhosis (Table S2). Data for advanced fibrosis or cirrhosis were reported separately. ELF cut-off values for studies reported using the original ELF algorithm were converted to current values by adding 10, as previously described.[8,16,27–29]

The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for the test were collated at all the reported thresholds. Where an optimal threshold was identified this was noted. The AUROC, where reported, was extracted as a summary measure of diagnostic accuracy capturing performance in a single value. Results between 0.9–1.0 were considered to be excellent, 0.8–0.9 as good and anything between 0.6 and 0.8 as fair/moderate. Where possible, the true positive, false positive, true negative, and false negative values were calculated based on the sensitivities, specificities, and biopsy results provided in the papers. Confidence intervals (CIs) for sensitivities and specificities, where not already provided, were calculated using these contingency table data (Table S3), and forest plots were generated for the diagnosis of advanced fibrosis and for the diagnosis of cirrhosis, using the most commonly used ELF threshold, respectively.

Assessment of methodological quality. The QUADAS-2 tool[30] was used to assess study quality and was adapted for this systematic review. The information for phase 1 and phase 2 of QUADAS-2 was collected in the data extraction tool and therefore was not duplicated. The questions in the amended QUADAS-2 tool and explanations of how risk of bias and concern over applicability were determined are shown in Table S4.

Results

Results of the search. After assessing 631 full-text articles, 36 were included (Fig. 1), of which 35 were identified from the



Abbreviations: CLD, chronic liver disease; HBV, hepatitis virus B; HCV, hepatitis virus C; NAFLD, non-alcoholic liver disease

Figure 1 Flow chart for the selection of articles. [Colour figure can be viewed at wileyonlinelibrary.com]

electronic search and one published before the ELF test was named was included as the index paper.

Characteristics of included studies

Study design and etiology. Eleven studies recruited patient with HCV,[31–41] 4 with HBV,[42–45] 9 with NAFLD,[16,27,28,46–51] 2 in patients with ALD,[52,53] and 10 with mixed etiology CLD.[8,54–62] Two of these studies[8,58] reported results for all participants collectively, as well as results for separate CLD etiologies. These results have been included in the etiology specific results.

Participants. Of the included studies, 31 were prospective,[8,16,27,28,31,32,34–40,42–44,46–57,60–62] and 5 retrospective.[33,38,41,45,58,59] 22 were conducted in Europe,[8,16,27,28,31–34,37,40,43,47,49–55,58,61,62] 2 in USA,[41,56] 6 in Asia,[36,42,44,45,48,60] 2 in South America,[35,39] 2 in Australia,[57,59] and 2 were international.[38,46] Sample sizes ranged from 38 to 3202.

The majority of the patients were male in 29 studies,[8,16,27,28,32–34,36,37,40–45,47,49–57,59–62] female in 6, and not reported in 1. The average age of patients (calculated by median or mean) ranged from 40 to 66 years.

The mean or median body mass index of participants was reported for 25 studies. The average body mass index was <25 in 8 studies,[32,34,39,40,42,44,45,60] 25–29 in 13 studies,[31,35,37,48–50,52,53,55–59] and >30 in 4 studies (in NAFLD).[16,28,47,51]

Reference standard. Half the studies did not state how the liver biopsy was obtained (n = 17).[8,16,28,34,38–41,45–47,56–59,61,62] The biopsy method was percutaneous in 16 studies,[31–33,35–37,42–44,48,50–52,60] and a mix of percutaneous or transjugular biopsies in 3.[27,54,55] The majority of studies used the METAVIR system for staging fibrosis (n = 17).[31,33–41,43,44,51,54,57–60] Other systems were Ishak,[32,61,62] Scheuer,[55] Batts and Ludwig,[42,45] the Clinical Research Network scoring system,[28,45–53] and the National Institute of Diabetes and Digestive and Kidney Diseases scoring system.[16] Dyvorne *et al*.[56] used a combination of METAVIR and Brunt, and Rosenberg *et al*.[8] used both Ishak and Scheuer.

The minimum required biopsy length was stated in 22 studies.[8,28,31–34,37,42–45,47,50–53,55–58,60,61] The remaining 14 studies reported average biopsy length or gave no information regarding biopsy standard. Thirteen studies set the minimum length of biopsies at ≥15 mm.[28,31,34,42–45,50,51,55–57,61] Seventeen studies stated the minimum number of portal tracts required for inclusion,[8,28,31,33–35,37,39,41,44,52,53,55,56,58,60,61] which was >6 in five studies.[8,39,41,52,56]

Index test. In 16 studies,[8,27,33,41–44,47,51–53,55,57,60–62] the ELF test and liver biopsy were performed on the same day. The time interval was up to a maximum of 6 months in 10 studies,[16,28,31,34,35,37–39,45,63] more than 6 months in 4 studies,[32,46,56,58] and unknown in 6 studies.[36,40,48,50,54,59]

A pre-defined ELF cut-off (manufacturer's recommendation or from earlier studies) was used in nine studies.[33,40,41,46,50–53,57,58]

Methodological quality of included studies. Only one study had a low risk of bias in all domains[51] while among the remaining 35 studies, 32 had low concern regarding applicability in all domains.[8,16,27,28,31,32,34–37,39–54,57–62] The HBV studies[42–45] were at low risk of bias except for the index test domain because they did not use pre-specified ELF test thresholds. Although only two studies evaluated the diagnostic accuracy of ELF in ALD, they were of high quality with concerns only in the "reference standard" domain for the Thiele *et al*.[52] study and in the patient selection domain for the Madsen *et al*.[53] study. In the latter, the flow and timing bias was unclear.

Only nine studies,[33,36,40,46,50–52,57,58] none of which included HBV patients, had a low risk of bias in the "index test" domain. Most studies did not use pre-specified ELF test thresholds, conferring a higher risk of bias in this domain for 24 studies[8,16,27,28,31,32,34,35,37–39,41–45,47–49,53–55,61,62] and unclear risk for three.[56,59,60] In the "patient selection" domain, only three studies were at high risk of bias (two HCV and one mixed)[8,31,32] but across all etiologies the level of concern was unclear in fourteen studies.[33,34,36,38,39,41,47–49,54,56,59,60,62] Fourteen studies[28,31,33–35,37,42–45,51,53,55,58] were at low risk in the "reference standard" domain, five[8,32,39,47,52] at high risk and in the remaining 17[9,16,27,36,40,41,46,48–50,54,56,57,59–62] the risk of bias was unclear.

The "flow and timing" domain had high risk in six studies[27,32,33,46,56,58] and unclear risk in six studies (Table 1).[16,36,48,50,53,54]

Findings

Hepatitis C virus

Advanced fibrosis. Eleven studies provided data on ELF in the diagnosis of advanced fibrosis in patients with HCV.[8,32,34–41,58]

The AUROCs for detecting advanced fibrosis in HCV patients ranged from 0.773 (95% CI 0.697–0.848)[8] to 0.98 (95% CI 0.93–1.00).[41]

Parkes *et al*.[38] reported multiple cut-offs for ELF ranging from 9.13 to 10.90. The optimal cut-off was chosen at 10.48, giving a sensitivity of 62% and specificity of 89%. Rosenberg *et al*.[8] reported cut-offs from 0.063 to 0.564 (using the original ELF algorithm, corresponding to 10.06 to 10.56), with the optimal threshold of 0.063 giving a sensitivity of 95% and specificity of 29%.

The range of cut-offs, including the optimal cut-offs for Parkes *et al*.[38] and Rosenberg *et al*.,[8] was from 9.33[34] to 10.59.[35] The sensitivity of ELF varied from 65%[40] to 100%.[41] The specificity ranged from 29%[8] to 99% (Table 2).[41]

Cirrhosis. Ten studies provided data on ELF in the diagnosis of cirrhosis in HCV patients.[31–37,39,41,58] The AUROCs ranged from 0.63 (95% CI 0.43–0.80)[33] to 0.99 (95% CI 0.93–1.00).[41]

The overall range of cut-offs was 8.1[33] to 11.27.[36] Martinez *et al*.[37] reported two cut-offs of 0.06 and 1.73 (using the original ELF algorithm, corresponding to 10.06 to 11.73), giving sensitivities of 90% and 52%, respectively, and specificities of 53% and 90%, respectively.

Table 1 Quality assessment of all the included papers, displayed in the QUADAS-2 format

Authors, date	Sample size	Study design	Quality assessment						
			Domain 1: patient selection		Domain 2: index test		Domain 3: Reference standard		Domain 4: flow and timing
			Bias	Applicability	Bias	Applicability	Bias	Applicability	Bias
Abdel-Hameed *et al.*, 2020[37]	98 (HCV monoinfection)	Retrospective	Unclear	Low	High	Low	Unclear	Low	Low
Catanzaro *et al.*, 2013[27]	162	Prospective	High	Low	High	Low	Low	Low	Low
Cobbold *et al.*, 2010[38]	80 recruited; 67 included	Prospective	High	Low	High	Low	High	Low	High
D'Ambrosio *et al.*, 2016[29]	38	Retrospective (post-hoc analysis)	Unclear	High	Low	Low	Low	Low	High
Fernandes *et al.*, 2015[31]	140 recruited; 120 included	Prospective	Low	Low	High	Low	Low	Low	Low
Fujita *et al.*, 2018[32]	122	Prospective	Unclear	Low	Low	Low	Unclear	Low	Unclear
Guechot *et al.*, 2012[30]	590 recruited 512 included	Prospective	Unclear	Low	High	Low	Low	Low	Low
Martinez *et al.*, 2011[33]	340	Cohort	Low	Low	High	Low	Low	Low	Low
Parkes *et al.*, 2011[34]	347 total C1: 87 C2: 173 C3: 87	Prospective	Unclear	Unclear	High	Low	Unclear	Low	Low
Ragazzo *et al.*, 2017[35]	250 recruited, 107 Included	Prospective	Unclear	Low	High	Low	High	Low	Low
Tanwar *et al.*, 2017[36]	108 recruited to PROFI-C trial; 80 included	Part of prospective RCT	Low	Low	Low	Low	Unclear	Low	Low
Heo *et al.*, 2018[41]	265	Retrospective	Low	Low	High	Low	Low	Low	Low
Kim *et al.*, 2012[38]	253 recruited, 170 included	Prospective	Low	Low	High	Low	Low	Low	Low
Trembling *et al.*, 2014[39]	224 recruited, 182 included	Prospective	Low	Low	High	Low	Low	Low	Low
Wong *et al.*, 2014[40]	238	Prospective	Low	Low	High	Low	Low	Low	Low
Anstee *et al.*, 2019[42]	3202	Prospective	Low	Low	Low	Low	Unclear	Low	High
Dvorak *et al.*, 2014[23]	112	Prospective	Low	Low	High	Low	Unclear	Low	High
Eddowes *et al.*, 2018[24]	54	Prospective	Low	Low	High	Low	Low	Low	Low
Guha *et al.*, 2008[22]	192	Prospective	Low	Low	High	Low	Unclear	Low	Unclear
Guillame *et al.*, 2019[43]	417	Prospective	Unclear	Low	High	Low	High	Low	Low
Inadomi *et al.*, 2020[44]	366	Prospective	Unclear	Low	High	Low	Unclear	Low	Unclear
Lykiardopoulos *et al.*, 2016[45]	158	Prospective	Unclear	Low	High	Low	Unclear	Low	Low
Miele *et al.*, 2017[46]	82	Prospective	Low	Low	Low	Low	Unclear	Low	Unclear
Staufer *et al.*, 2019[47]	186	Prospective	Low	Low	Low	Low	Low	Low	Low
Madsen *et al.*, 2020[49]	266	Prospective	Low	Low	High	Low	Low	Low	Unclear
Thiele *et al.*, 2018[48]	289	Prospective	Low	Low	Low	Low	High	Low	Low
Agrawal *et al.*, 2016[50]	115	Prospective	Unclear	Low	High	Low	Unclear	Low	Unclear
Crespo *et al.*, 2012[51]	146 Non-transplant = 59 Transplant = 87	Prospective	Low	High	High	Low	Low	Low	Low
Dyvorne *et al.*, 2016[52]	60	Prospective	Unclear	Low	Unclear	Low	Unclear	Unclear	High
Fagan *et al.*, 2015[53]	536 patients recruited 318 included	Prospective	Low	Low	Low	Low	Unclear	Low	Low

(Continues)

Table 1 (Continued)

Authors, date	Sample size	Study design	Quality assessment						
			Domain 1: patient selection		Domain 2: index test		Domain 3: Reference standard		Domain 4: flow and timing
			Bias	Applicability	Bias	Applicability	Bias	Applicability	Bias
Friedrich-Rust *et al.*, 2010[54]	74	Retrospective	Low	Low	Low	Low	Low	Low	High
Irvine *et al.*, 2016[55]	432	Retrospective	Unclear	Low	Unclear	Low	Unclear	Low	Low
Lee *et al.*, 2010[56]	312 recruited 280 included (estimation group = 121; validation group = 159)	Prospective	Unclear	Low	Unclear	Low	Unclear	Low	Low
Rosenberg *et al.*, 2004[8]	1021 recruited (test = 400, validation = 521)	Prospective	High	Low	High	Low	High	Low	Low
Stasi *et al.*, 2019[57]	143 recruited 115 included	Prospective	Low	Low	High	Low	Unclear	Low	Low
Wahl *et al.*, 2012[58]	102	Prospective	Unclear	Low	High	Low	Unclear	Low	Low

Abbreviations: AIH, autoimmune hepatitis; ALD, alcohol liver disease; CC, cryptogenetic cirrhosis; NAFLD, non-alcoholic liver disease; HBV, hepatitis virus B; HCV, hepatitis virus C; PBC, primary biliary cholangitis; PSC, primary sclerosing cholangitis.

The sensitivity ranged from 7%[41] to 100%.[39] The specificity ranged from 53%[37] to 100%.[39]

Hepatitis B virus

Advanced fibrosis. All four studies looking at HBV provided data on the diagnostic accuracy of ELF in advanced fibrosis.[42–45] The AUROCs ranged from 0.69 (95% CI 0.63–0.75)[43] to 0.86 (95% CI 0.81–0.92).[42]

Trembling *et al.*[43] reported multiple cut-offs between 8.02 and 10.41; the sensitivities using these cut-offs ranged from 45% to 96%, and specificities ranged from 17% to 95%.

Kim *et al.*[42] used a cut-off of 9.40 providing 84% sensitivity and 78% specificity. Wong *et al.*[44] reported sensitivity of 62% and specificity of 66% using a cut-off of 9.8. Heo *et al.*[45] reported a cut-off of 9.8 providing the maximum sum of sensitivity and specificity.

Cirrhosis. All four studies looking at HBV provided data on the diagnostic accuracy of ELF in cirrhosis.[42–45] The AUROCs ranged from 0.706 0.68 (95% CI 0.61–0.75)[44] to 0.86 (95% CI 0.81–0.92).[42]

Trembling *et al.*[43] reported multiple cut-offs between 8.61 and 10.68; the sensitivities using these cut-offs ranged from 44% to 94%, and specificities ranged from 39% to 95%. Kim *et al.*[42] reported a sensitivity of 70% and a specificity of 79% using a cut-off of 10.10. Wong *et al.*[44] used a cut-off of 9.5 giving a sensitivity of 78% and specificity of 47%.

Heo *et al.*[45] reported a cut-off of 9.5 as the one providing the maximum sum of sensitivity and specificity. This threshold is lower than the threshold quoted in the same study as giving the maximum sensitivity and specificity for the diagnosis of advanced fibrosis.

Non-alcoholic fatty liver disease

Advanced fibrosis. The AUROCs for detecting advanced fibrosis in NAFLD patients ranged from 0.78 (0.70–0.89)[49] to 0.97 (no CI reported).[27]

Dvorak *et al.*[27] looked at two different thresholds and for the optimal threshold of −3.37 (corresponding to 6.63 after the addition of 10) found that the ELF test had a sensitivity of 90% and a specificity of 97%. Guha *et al.*[16] found 0.3576 (corresponding to 9.64) to be the most optimal threshold, with a sensitivity of 80% and a specificity of 90%. Anstee *et al.*[46] reported a sensitivity of 20% and a specificity of 98% with a cut-off of 11.3. Guillaume *et al.*[47] reported an optimal sensitivity of 73% and specificity of 72%. Inadomi *et al.*[48] reported results for two different cohorts finding 10.38 to be the optimal threshold with a sensitivity of 63% and 70% and a specificity of 79% and 81% according to the cohort. Miele *et al.*[50] and Staufer *et al.*[51] used a threshold of 9.8 reporting a sensitivity of 87%% and 72% and specificity of 93% and 90%, respectively.

Rosenberg *et al.*[8] included NAFLD patients in their mixed etiology study and reported the results for these patients separately. The optimal threshold for detecting advanced fibrosis was identified as 0.375 (using original ELF algorithm, corresponding to 9.625) which provided a sensitivity of 89% and a sensitivity of 96%.

Cirrhosis. Only 2 studies reported the AUROCs for detecting cirrhosis in NAFLD patients which were 0.852 ± 0.040 in Guillaume *et al.*[47] and 0.92 (0.88–0.97) in Staufer *et al.*[51] No study

Table 2 Summary of the data extracted from the included papers

Authors, date	Etiology	Target condition	Threshold	Sensitivity (95% CI)	Specificity (95% CI)	PPV	NPV	AUROC (95% CI)
Agrawal *et al.*, 2016[50]	Mixed (pooled)	Advanced fibrosis	9.29	60.0%	69.7%	—	—	0.707 (0.550–0.864)
		Cirrhosis	10.12	100.0%	84.1%	—	—	0.926 (0.843–1.000)
Crespo *et al.*, 2012[51]	Mixed (pooled)	Advanced fibrosis	—	—	—	—	—	—
		Cirrhosis	Non-transplant = 10.4 Transplant = 10.3 Training = 10.3	Non-transplant = 93% Transplant = 78% Training = 92%	Non-transplant = 79% Transplant = 72% Training = 72%	Non-transplant = 61% Transplant = 24% Training = 36%	Non-transplant = 97% Transplant = 96% Training = 98%	Non-transplant (n = 59) 0.894 Transplant (n = 87) 0.834
Dyvorne *et al.*, 2016[52]	Mixed (pooled)	Advanced fibrosis	—	—	—	—	—	0.63
		Cirrhosis	—	—	—	—	—	—
Fagan *et al.*, 2015[53]	Mixed (pooled)	Advanced fibrosis	Manufacturer cut-off = 9.8 Modified cut-off = 9.7	Manufacturer = 74.4% Modified = 76.9%	Manufacturer = 92.4% Modified = 90.0%	Manufacturer = 75.3%	Manufacturer = 92.1%	Modified = 0.91 (0.88–0.95)
		Cirrhosis	Modified cut-off = 10.2	Modified = 68.3%	Modified = 90.0%	—	—	Modified = 0.90 (0.84–0.95)
Friedrich-Rust *et al.*, 2010[54]	Mixed (pooled)	Advanced fibrosis	10.22	74%	70%	64%	79%	0.79 (0.67–0.91)
		Cirrhosis	10.31	91%	62%	29%	98%	0.92 (0.83–1.00)
Irvine *et al.*, 2016[55]	Mixed (pooled)	Advanced fibrosis	—	81.1%	80%	51.8%	94.1	0.898
		Cirrhosis	—	—	—	—	—	—
Lee *et al.*, 2010[56]	Mixed (pooled)	Advanced fibrosis	—	—	—	—	—	—
		Cirrhosis	—	—	—	—	—	Estimation group (n = 121) 0.790 (0.663–0.917) Validation group (n = 159) 0.775 (0.692–0.859) Combined (n = 280) 0.698 (0.611–0.776)
Rosenberg et al., 2004[8]	Mixed (pooled)	Advanced fibrosis	Validation group (n = 521) 0.063 0.102 0.130 0.179 0.238 0.273 0.358 0.457 0.507 0.826	Validation group (n = 521) 95% 90% 85% 80% 69% 64% 54% 47% 44% 19%	Validation group (n = 521) 24% 41% 53% 67% 80% 85% 90% 95% 96% 99%	Validation group (n = 521) 31% 35% 40% 46% 55% 60% 65% 75% 80% 90%	Validation group (n = 521) 93% 92% 91% 90% 88% 87% 84% 83% 83% 77%	Test and validation groups (n = 1021) 0.804 (0.757–0.850)
		Cirrhosis	0.025	90.7%	69.2%	—	—	Test and validation groups (n = 1,021) 0.887 (0.837–0.937)

(Continues)

Table 2 (Continued)

Authors, date	Etiology	Target condition	Threshold	Sensitivity (95% CI)	Specificity (95% CI)	PPV	NPV	AUROC (95% CI)
Stasi *et al.*, 2019[57]	Mixed (pooled)	Advanced fibrosis	12	30.43	97.75	—	—	0.785 (0.702–0.854)
		Cirrhosis	12	46.15	96.97	—	—	0.880 (0.821–0.932)
Wahl *et al.*, 2012[58]	Mixed (pooled)	Advanced fibrosis	—	—	—	—	—	—
		Cirrhosis	9.39	100%	77%	—	—	0.93 (0.88–0.99)
Abdel-Hameed *et al.*, 2020[37]	HCV	Advanced fibrosis	9.49	100%	90%	81%	100%	0.98 (0.93–1.00)
			9.62	97%	94%	88%	98%	
			9.8	73%	99%	96%	89%	
		Cirrhosis	9.49	100%	82%	48%	100%	0.99 (0.93–1.00)
			9.8	93%	88%	57%	99%	
			10.17	71%	98%	83%	95%	
			10.41	64%	99%	90%	94%	
			11.14	7%	100%	100%	87%	
Catanzaro *et al.*, 2013[27]	HCV	Advanced fibrosis	—	—	—	—	—	—
		Cirrhosis	9.30	79%	91%	76%	92%	0.94 (0.88–0.96)
Cobbold *et al.*, 2010[28]	HCV	Advanced fibrosis	8.75	84	70	78	78	0.82 (0.73–0.92)
		Cirrhosis	9.4	93%	79%	54%	98%	0.91 (0.82–1.00)
D'Ambrosio *et al.*, 2016[29]	HCV	Advanced fibrosis	—	—	—	—	—	—
		Cirrhosis	8.1	60%	74%	54%	78%	0.63 (0.43–0.80)
			9.8 (n = 2)	90%	10%	34%	67%	
Fernandes *et al.*, 2015[31]	HCV	Advanced fibrosis	10.59	74%	90%	58%	95%	0.82 (0.74–0.88)
		Cirrhosis	10.44	88%	78%	22%	99%	0.78 (0.70–0.85)
Friedrich-Rust *et al.*, 2010[54]	HCV (mixed)	Advanced fibrosis	10.22	82%	74%	74%	82%	—
		Cirrhosis	10.31	89%	63%	44%	94%	—
Fujita *et al.*, 2018[32]	HCV	Advanced fibrosis	9.97	85.7%	61.5%	—	—	0.81 (0.77–0.91)
			11.27	64.3%	87.9%			
		Cirrhosis	—	—	—	—	—	—
Guechot *et al.*, 2012[30]	HCV	Advanced fibrosis	9.33	90%	63%	73%	85%	0.82 (0.78–0.86)
		Cirrhosis	9.35	83%	75%	44%	95%	0.85 (0.81–0.90)
Martinez *et al.*, 2011[33]	HCV	Advanced fibrosis	—	—	—	—	—	0.83 (0.79–0.87)
		Cirrhosis	0.06	90%	53%	52%	90%	0.82 (0.78–0.87)
			1.73	52%	90%	76%	77%	
Parkes *et al.*, 2011[34]	HCV	Advanced fibrosis	9.13	95%	44%	44%	95%	0.85 (0.80–0.89)
			9.39	90%	55%	48%	92%	
			9.59	85%	63%	52%	90%	
			10.22	70%	85%	68%	86%	
			10.48	62%	89%	73%	83%	
			10.90	54%	95%	82%	81%	
		Cirrhosis	—	—	—	—	—	—
Ragazzo *et al.*, 2017[35]	HCV	Advanced fibrosis	9.47	83%	69%	37%	95%	0.82 (0.71–0.93)
		Cirrhosis	11.0	20%	100%	100%	94%	0.94 (0.91–1)

(Continues)

Table 2 (Continued)

Authors, date	Etiology	Target condition	Threshold	Sensitivity (95% CI)	Specificity (95% CI)	PPV	NPV	AUROC (95% CI)
Rosenberg *et al.*, 2004[8]	HCV (mixed)	Advanced fibrosis	0.063	95%	29%	27.7%	94.9%	Validation group (n = 521)
			0.067	90%	31%	27.5%	92.3%	0.773 (0.697–0.848)
			0.090	85%	43%	29.9%	91.1%	
			0.126	80%	58%	35.2%	91.0%	
			0.190	63%	80%	47.9%	88.5%	
			0.219	52%	85%	50.0%	86.2%	
			0.268	47%	90%	57.8%	85.6%	
			0.426	38%	95%	70.0%	84.3%	
			0.564	30%	99%	89.5%	83.3%	
		Cirrhosis	—	—	—	—	—	—
Tanwar *et al.*, 2017[36]	HCV	Advanced fibrosis	9.59	65%	82%	71%	78%	0.82 (0.72–0.92)
		Cirrhosis	—	—	—	—	—	0.89 (0.79–1.00)
Heo *et al.*, 2018[41]	HBV	Advanced fibrosis	8.4	95.0%	34.6%	71.7%	80.0%	0.703 (0.638–0.762)
			9.8	53.2%	71.6%	76.5%	46.8%	
			10.8	24.8%	92.6%	85.4%	41.4%	
		Cirrhosis	8.8	86.4%	38.5%	61.4%	71.4%	0.706 (0.642–0.765)
			9.5	66.1%	58.7%	64.5%	60.4%	
			11.1	21.2%	91.3%	73.5%	50.5%	
Kim *et al.*, 2012[38]	HBV	Advanced fibrosis	9.40	84%	78%	79%	83%	0.86 (0.81–0.92)
		Cirrhosis	10.10	70%	79%	56%	87%	0.86 (0.81–0.92)
Trembling *et al.*, 2014[39]	HBV	Advanced fibrosis	8.02	96%	17%	40%	86%	0.8 (0.73–0.87)
			8.45	93%	41%	48%	90%	
			8.96	85%	56%	53%	86%	
			9.39	73%	70%	58%	82%	
			9.88	60%	83%	67%	78%	
			10.41	45%	95%	83%	75%	
		Cirrhosis	8.61	94%	39%	28%	97%	0.83 (0.76–0.90)
			9.43	72%	64%	34%	90%	
			9.66	69%	72%	38%	90%	
			9.99	67%	81%	47%	91%	
			10.34	61%	87%	54%	90%	
			10.68	44%	95%	70%	87%	
Wong *et al.*, 2014[40]	HBV	Advanced fibrosis	9.8	62%	66%	55%	72%	0.69 (0.63–0.75)
		Cirrhosis	9.5	78%	47%	31%	88%	0.68 (0.61–0.75)
Anstee *et al.*, 2019[42]	NAFLD	Advanced fibrosis	11.3	20% (19–22)	98% (96–99)	95% (93–97)	33% (32–35)	0.80 (0.80–0.80)
		Cirrhosis	—	—	—	—	—	—
Dvorak *et al.*, 2014[23]	NAFLD	Advanced fibrosis	−3.37	90%	97%	—	—	0.97
			−3.39	93%	93%	—	—	
		Cirrhosis	—	—	—	—	—	—
Eddowes *et al.*, 2018[24]	NAFLD	Advanced fibrosis	—	—	—	—	—	0.80 (0.68–0.93)
		Cirrhosis	—	—	—	—	—	—

(Continues)

Table 2 (Continued)

Authors, date	Etiology	Target condition	Threshold	Sensitivity (95% CI)	Specificity (95% CI)	PPV	NPV	AUROC (95% CI)
Guha *et al.*, 2008[22]	NAFLD	Advanced fibrosis	−1.12413	100%	12%	26%	100%	0.90 (0.84–0.96)
			−0.7121	98%	42%	34%	98%	
			−0.4184	96%	57%	41%	98%	
			−0.1068	90%	75%	52%	96%	
			0.3576	80%	90%	71%	94%	
			0.8139	62%	95%	78%	89%	
			1.6454	29%	99%	87%	82%	
			2.2858	16%	100%	100%	80%	
		Cirrhosis	—	—	—	—	—	—
Guillaume *et al.*, 2019[43]	NAFLD	Advanced fibrosis	9.3	73%	72%	64%	90%	0.793 ± 0.022
			10.0	47%	90%	75%	72%	
		Cirrhosis	—	—	—	—	—	0.852 ± 0.040
Inadomi *et al.*, 2020[44]	NAFLD	Advanced fibrosis	10.38	70%	81%	68%	82%	0.806 (0.741–0.871)
		(training set)	9.34	90%	31%	43%	85%	
			10.83	52%	91%	76%	77%	
		Advanced fibrosis	10.38	63%	79%	55%	84%	0.812 (0.752–0.872)
		(validation set)	9.34	90%	38%	38%	90%	
			10.83	53%	86%	61%	81%	
Lykiardopoulos *et al.*, 2016[45]	NAFLD	Advanced fibrosis	—	—	—	—	—	0.78 (0.70–0.89)
		Cirrhosis	—	—	—	—	—	—
Miele *et al.*, 2017[46]	NAFLD	Advanced fibrosis	9.8	87% (CI 0.69–1.04)	93% (86–99)	72%	97%	0.948 (0.88–1.00)
		Cirrhosis	—	—	—	—	—	—
Rosenberg *et al.*, 2004[8]	NAFLD (mixed)	Advanced fibrosis	0.375	89%	96%	80%	98%	Validation group
			0.462	78%	98%	87%	96%	0.870 (0.666–1.000)
		Cirrhosis	—	—	—	—	—	—
Staufer *et al.*, 2019[47]	NAFLD	Advanced fibrosis	9.8	72%	90%	70%	90%	0.90 (0.85–0.95)
		Cirrhosis	—	—	—	—	—	0.92 (0.88–0.97)
Madsen *et al.*, 2020[49]	ALD	Advanced fibrosis	10.5	77% (64–87)	90% (85–94)	71% (58–81)	93% (88–96)	0.92 (0.88–0.96)
		Cirrhosis	10.1	93% (82–99)	80% (74–85)	48% (37–59)	98% (95–100)	0.93 (0.90–0.97)
Rosenberg *et al.*, 2004[8]	ALD (mixed)	Advanced fibrosis	0.087	100%	16.7%	75%	100%	Validation group
			0.431	93.3%	100%	100%	85.7%	0.944 (0.836–1.000)
		Cirrhosis	—	—	—	—	—	—
Thiele *et al.*, 2018[48]	ALD	Advanced fibrosis	9.8	89% (79–96)	78% (72–83)	54% (44–64)	96% (92–98)	0.92 (0.89–0.96)
			10.51	79% (67–88)	91% (86–94)	71% (59–81)	94% (89–96)	
		Cirrhosis	—	—	—	—	—	0.94 (0.91–0.97)

Abbreviations: ALD, alcohol liver disease; AUROC, area under the receiver operator curve; ELF, enhanced liver fibrosis; HBV, hepatitis virus B; HCV, hepatitis virus C; NAFLD, non-alcoholic liver disease; NPV, negative predictive value; PPV, positive predictive value.

The papers have been organized according to chronic liver disease etiology, with all reported ELF test thresholds for the diagnosis of advanced fibrosis or cirrhosis.

reported sensitivity and specificity for detecting cirrhosis in NAFLD patients.

Alcohol liver disease

Advanced fibrosis. Only three studies reported the diagnostic accuracy of ELF in ALD patients.[8,52,53] The AUROC was excellent ranging from 0.92 (0.89–0.96) in the Thiele *et al*. study[52] and in the Madsen *et al*. study (0.88–0.96) to 0.944 (0.836–1.000).[8] Using the original ELF algorithm, Rosenberg *et al*.[8] identified 0.087 as the optimal threshold (corresponding to 9.913), which provided a sensitivity of 100%, but a specificity of 16.7% (PPV 75%, NPV 100%). In the Thiele *et al*.[52] study, the manufacturer's threshold (9.8) was used as well as the threshold recommended by the NICE guidelines for NAFLD (10.51)[64] with a sensitivity of 89% and 78% and a specificity of 78% and 91%, respectively.

Cirrhosis. Two studies[52,53] assessed the diagnostic accuracy of ELF at detecting cirrhosis reporting an excellent AUROC ranging from 0.93 (0.90–0.97) to 0.94 (0.91–0.97). Madsen *et al*.[53] reported a sensitivity of 93% and a specificity of 80% using a threshold of 10.1.

Mixed

Advanced fibrosis. Seven of the 10 studies enrolling patients with mixed etiology CLD assessed the diagnostic accuracy of the ELF test for detecting advanced fibrosis.[8,54,56–59,61]

In original ELF study, enrolling a mixed cohort of 1021 patients, the reported AUROC was 0.804 (0.757–0.850). and the optimal cut-off was identified at 0.102 (corresponding to 9.89, sensitivity of 90% and specificity of 41%).[8]

The AUROCs reported in the included studies ranged widely from 0.63 (no CI)[56] to 0.91 (0.88–0.95).[57] Dyvorne *et al*.'s[56] study was primarily in HCV patients (81.6%) and with a small sample size (n = 60). In comparison, Fagan *et al.et al*.[57] enrolled 318 patients, 60.2% of whom had HCV.

Agrawal *et al*.[54] enrolled 115 patients (55.7% NAFLD) and reported an AUROC of 0.707 with a relatively low sensitivity and specificity of 60% and 69.7% respectively. Friedrich-Rust *et al*.[58] had reported a higher sensitivity of 74% sensitivity and 70% specificity, using a higher cut-off of 10.22 for advanced fibrosis. Stasi *et al*.[61] used the highest cut-off for advanced fibrosis (12) with an AUROC of 0.785 (0.702–0.854), a low sensitivity (30.43%) but the highest specificity (97.75%). Irvine *et al*.[59] reported an AUROC of 0.898, sensitivity of 81.1%, and specificity of 80%, but they did not specify the cut-off used.

Cirrhosis. Eight of the studies enrolling patients with a mix of causes of CLD reported the accuracy of the ELF test for detecting cirrhosis.[8,54,55,57,58,60–62] All of the AUROCs reported were above 0.80,[8,54–59,61,62] with the exception of one article, conducted in 280 patients with viral hepatitis, which reported an AUROC of 0.698 (no sensitivity or specificity reported).[60]

Rosenberg *et al*.[8] reported an AUROC of 0.887 (0.837–0.937) with a sensitivity of 90.7% and specificity of 69.2% for a cut-off value of 0.025 (corresponding to 9.975). Wahl *et al*.[62] reported an AUROC of 0.93 (0.88–0.99, sensitivity 100%, and specificity 77%) using a cut-off value of 9.39. Similarly, Friedrich-Rust *et al*.[58] reported an AUROC of 0.92 (0.83–1.00) (cut-off 10.31), and Agrawal *et al*.[54] had an AUROC of 0.926 (0.843–1.00) from a cut-off of 10.12. Fagan *et al*.[57] also had an AUROC of 0.9 from a cut-off of 10.2. In Fagan *et al*.[57] and Friedrich Rust *et al*.,[58] the majority of patients had HCV; in contrast, Agrawal *et al*.[54] had only 21% viral hepatitis patients. Stasi *et al*.[61] reported an AUROC of 0.880 (0.821–0.932, sensitivity 46.15%, and specificity 96.97%) using a cut-off of 12, which they also applied to detect advanced fibrosis.

Crespo *et al*.[55] looked at two cohorts of patients; in the non-transplant cohort, they found an AUROC of 0.894 and in the transplant patients an AUROC of 0.834. The non-transplant group had 41% HCV patients, compared with 72% HCV in the transplant group.

Forest plots (Fig. 2) revealed good sensitivity and specificity for ELF for detecting advanced fibrosis (F3, ELF = 9.8 ± 0.1) and for cirrhosis (F4, ELF = 10.3 ± 0.1) except for studies in HBV where the performance of ELF was consistently worse than in other etiologies.

Discussion

Summary of main results. This systematic review identified 36 studies assessing the accuracy of the ELF test for detecting advanced fibrosis or cirrhosis in patients with HCV, HBV, NAFLD, or ALD.[8,16,27,28,31–62] Eleven HCV studies,[31–41] 4 HBV studies,[42–45] 9 NAFLD studies,[16,27,28,46–51] 2 ALD studies,[52,53] and 10 mixed etiology CLD studies[8,54–62] were included.

In patients infected with HCV, there is a good quantity of evidence showing fair to excellent performance of the ELF test in detecting advanced fibrosis with slightly better performance for detecting cirrhosis. The quality of these studies is mixed, and few employed predetermined thresholds. However, the studies suggest that ELF is of use in assessing fibrosis in chronic hepatitis C.

While the numbers of studies in HBV patients are fewer, they are of higher quality, with the ELF test performing similarly well with AUROCs ranging from 0.69[44] to 0.86[42] for advanced fibrosis and cirrhosis. The study by Kim *et al*.[42] investigated a homogenous cohort of Asian patients, so could be considered less applicable to other ethnicities. The good AUROCs but poor sensitivity around the 9.8 and 10.3 ELF thresholds, as shown in Figure 2, suggest that disease-specific thresholds may be required in HBV infection.

This review found that in NAFLD the ELF test performs very well for detecting advanced fibrosis and cirrhosis. The findings of this review are aligned with the evidence used to make the recommendation in the NICE guideline on the assessment and management of NAFLD.[64]

The recent systematic review of Vali *et al*.[25] presented evidence of good diagnostic performance of ELF in the detection of advanced fibrosis and cirrhosis when used as intended, for the investigation of fibrosis in patients with NAFLD. While their

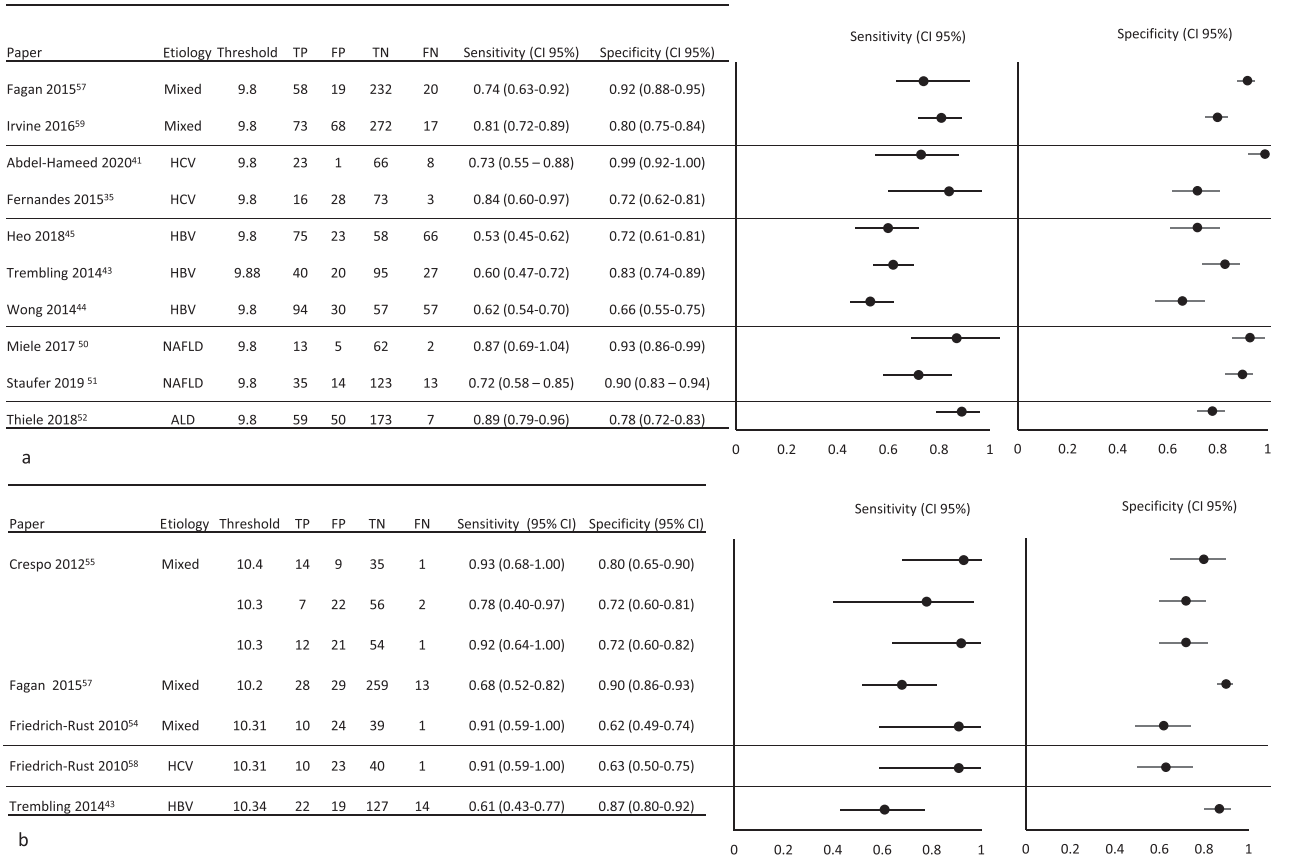


Figure 2 Forest plots of sensitivity and specificity of ELF for the detection of (a) advanced fibrosis and (b) cirrhosis. ALD, alcohol liver disease; CI, confidence interval; CLD, chronic liver disease; FN, false negative; FP, false positive; HBV, hepatitis B virus; HCV, hepatitis C virus; NAFLD, non-alcoholic liver disease; TP, true positive; TN, true negative.

modeling suggested that, like all other NITs, ELF may perform less well in low prevalence settings, the present systematic review focused on fibrosis assessment in patients with known or suspected CLD, the context in which the ELF test has been established and for which its use has regulatory approval.

In ALD patients, there is a small amount of evidence of high quality showing an excellent performance of ELF test with AUROCs of 0.92[52,53] and 0.944[8] for advanced fibrosis, and 0.92[53] and 0.94[52] for cirrhosis.

It is difficult to draw firm conclusions from the 10 studies[8,54–62] that recruited patients with differing etiologies given the heterogeneity and varying methodological quality. However, these mixed studies provided evidence that ELF can be used with good results across different etiologies.

Strengths and weaknesses of review. This systematic review is the first to bring together the evidence on the diagnostic accuracy of the ELF test for detecting advanced fibrosis and cirrhosis in a range of common CLD etiologies. The review has followed the guidelines for DTA systematic reviews and has used the QUADAS-2 tool[30] to assess the quality of the included studies. Our review did not include a meta-analysis due to the considerable study heterogeneity but does add weight and support to previous findings that the ELF test can be used across a range of CLD etiologies.[24] Furthermore, we were able to consider the evidence for the use of ELF in different disease etiologies, as well as in a mixed group of patients.

In addition to the clinical heterogeneity, there was significant methodological heterogeneity in the conduct of the index test and reference standard. Although the automation of the ELF test limits the potential to introduce bias, several studies scored highly for risk of bias in the index test domain because they did not use predetermined cut-off values for fibrosis detection. Several studies explored performance at multiple ELF thresholds and selected values providing the maximum sensitivity and specificity. This may overestimate ELF test performance but reflects the slow emergence of consensus around the appropriate cut-offs for different disease etiologies. Some studies are not easily comparable because they used different versions of the ELF algorithm to calculate a score. These differences arise from the addition of 10 to the original ELF scores to generate only positive values, and then subsequently a change in the algorithm used to calculate the ELF score due to the use of different auto-analyzers (Immuno-1 and Advia Centaur).[65] However, the manufacturer has demonstrated equivalence in measurement of analytes across the range

of ELF values between the Immuno-1 and Advia algorithms (ELF Test Instructions for Use, Siemens Healthineers, Tarrytown, New York, USA).

One of the major sources of methodological heterogeneity in the included studies was the quality of the biopsy reference standard. While some studies specified strict criteria for length and/or number of portal tracts, others provided no information on biopsy. The influence of biopsy length on staging accuracy has been studied in depth, and a biopsy length of at least 15 mm is commonly accepted for reliable staging.[4] Additionally, the use of a single pathologist has been recommended to eliminate inter-observer variation. Wong *et al*.[44] suggested that due to these problems with biopsy sampling and inter/intra-observer variability, a perfect non-invasive marker can only achieve an AUROC of approximately 90% when compared with a reference biopsy. Selection bias may be present in this review because we only included studies in which patients had undergone liver biopsy, and so the patients enrolled in these studies are likely to have had a higher prevalence of significant fibrosis or cirrhosis than other patients being investigated for CLD but not subjected to liver biopsy and even greater than in the general population. This resulting spectrum bias means that the performance of ELF in these studies is unlikely to reflect its performance in primary care or community settings where the prevalence of fibrosis is lower than in secondary care. This will result in a lower PPV and a higher NPV than in secondary care, which means ELF will be a better test for excluding advanced CLD but perform less well in identifying cases of advanced fibrosis or cirrhosis in a general population. In accordance with this, none of the studies recommended use of the ELF test to screen the general population.

Ultimately, the best way to validate ELF would be further study its prognostic performance in predicting long-term liver-related morbidity and mortality,[9,66] as has been evaluated in some other NITs.[67,68]

Further limitations of this review are that the diagnostic accuracy of the ELF test for detecting mild or moderate fibrosis was not considered, nor was the diagnostic accuracy of the ELF test in less common causes of CLD reviewed. Although the diagnostic accuracy of other NITs was not evaluated in this study, similar systematic reviews have been conducted in other NITs.[69–71] Finally, while studies have demonstrated the utility of the ELF test in pediatric CLD, especially NAFLD,[29] this systematic review only evaluated the ELF test in adult patients.

Conclusion

In summary, the ELF test showed good diagnostic performance in detecting advanced fibrosis in patients with viral hepatitis and excellent performance in NAFLD and ALD. There is also evidence of good diagnostic performance for detecting cirrhosis in patients with viral hepatitis and excellent performance in patients with ALD. The quality of studies in HBV and ALD patients was very high, but more variable for HCV and NAFLD patients.

This review suggests that the ELF test could offer an alternative to biopsy for assessing liver fibrosis in viral hepatitis, NAFLD, and ALD. However, the included studies were significantly heterogeneous, and further comparative studies of high methodological quality are desirable. The ELF test also offers other benefits such as lack of operator variability, excellent pre-analytical and analytical performance, and the very low failure rate, which is restricted to situations where a blood sample cannot be obtained. Furthermore, the automation of the ELF test means that it can be used efficiently to test large numbers of patients. Although beyond the scope of this review, the ELF test may offer the advantage over liver biopsy of dynamic monitoring of fibrosis progression or regression, for example, following treatment directed at underlying causes.

Meta-analysis of the diagnostic accuracy of each disease etiology should also be considered in future studies.

Acknowledgments

The authors acknowledge the support of National Institute for Health Research.

References

1 Williams R, Aspinall R, Bellis M *et al*. Addressing liver disease in the UK: a blueprint for attaining excellence in health care and reducing premature mortality from lifestyle issues of excess consumption of alcohol, obesity, and viral hepatitis. *Lancet* 2014; **384**: 1953–97.
2 Lucero C, Brown RS Jr. Noninvasive measures of liver fibrosis and severity of liver disease. *J. Gastroenterol. Hepat. (Australia).* 2016; **12**: 33–40.
3 Bedossa P, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. *Hepatology* 2003; **38**: 1449–57.
4 Regev A, Berho M, Jeffers LJ *et al*. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. *Am. J. Gastroenterol.* 2002; **97**: 2614–8.
5 Mederacke I, Wursthorn K, Kirschner J *et al*. Food intake increases liver stiffness in patients with chronic or resolved hepatitis C virus infection. *Liver Int.* 2009; **29**: 1500–6.
6 Castera L, Foucher J, Bernard PH *et al*. Pitfalls of liver stiffness measurement: a 5-year prospective study of 13,369 examinations. *Hepatology* 2010; **51**: 828–35.
7 Myers R, Pomier-Layrargues G, Kirsch R *et al*. Feasibility and diagnostic performance of the FibroScan XL probe for liver stiffness measurement in overweight and obese patients. *Hepatology* 2012; **55**: 199–208.
8 Rosenberg WM, Voelker M, Thiel R *et al*. Serum markers detect the presence of liver fibrosis: a cohort study. *Gastroenterology* 2004; **127**: 1704–13.
9 Parkes J, Roderick P, Harris S *et al*. Enhanced liver fibrosis test can predict clinical outcomes in patients with chronic liver disease. *Gut* 2010; **59**: 1245–51.
10 Kennedy OJ, Primary Care & Population Sciences FoM, University of Southampton, Southampton General Hospital, Southampton, UK, Parkes J *et al*. The Enhanced Liver Fibrosis (ELF) Panel: analyte stability under common sample storage conditions used in clinical practice. *J. Appl. Lab. Med* 2020; **1**: 720–8.
11 Selby P, Banks RE, Gregory W *et al*. Methods for the evaluation of biomarkers in patients with kidney and liver diseases: multicentre research programme including ELUCIDATE RCT. *Programme Grants Appl. Res.* 2018; **6**: 1–528.
12 Gaia S, Carenzi S, Barilli A *et al*. Reliability of transient elastography for the detection of fibrosis in non-alcoholic fatty liver disease and chronic viral hepatitis. *J. Hepatol.* 2011; **54**: 64–71.
13 Macaluso F, Maida M, Cammà C *et al*. Steatosis affects the performance of liver stiffness measurement for fibrosis assessment in patients with genotype 1 chronic hepatitis C. *J. Hepatol.* 2014; **61**: 523–9.

14 Boursier J, de Ledinghen V, Sturm N *et al*. Precise evaluation of liver histology by computerized morphometry shows that steatosis influences liver stiffness measured by transient elastography in chronic hepatitis C. *J. Gastroenterol.* 2014; **49**: 527–37.
15 López I, Aroca F, Bernal M *et al*. Utility of the ELF test for detecting steatohepatitis in morbid obese patients with suspicion of nonalcoholic fatty liver disease. *Obes. Surg.* 2017; **27**: 2347–53.
16 Guha IN, Parkes J, Roderick P *et al*. Noninvasive markers of fibrosis in nonalcoholic fatty liver disease: validating the European Liver Fibrosis Panel and exploring simple markers. *Hepatology* 2008; **47**: 455–60.
17 Abignano G, Blagojevic J, Bissell LA *et al*. European multicentre study validates enhanced liver fibrosis test as biomarker of fibrosis in systemic sclerosis. *Rheumatology (Oxford)* 2019; **58**: 254–9.
18 Crossan C, Tsochatzis EA, Longworth L *et al*. Cost-effectiveness of non-invasive methods for assessment and monitoring of liver fibrosis and cirrhosis in patients with chronic liver disease: systematic review and economic evaluation. *Health Technol. Assess.* 2015; **19**: 1–409.
19 Shiha G, Ibrahim A, Helmy A *et al*. Asian-Pacific Association for the Study of the Liver (APASL) consensus guidelines on invasive and non-invasive assessment of hepatic fibrosis: a 2016 update. *Hepatol Int.* 2017; **11**: 1–30.
20 European Association for the Study of the Liver. EASL Clinical Practice Guidelines: the diagnosis and management of patients with primary biliary cholangitis. *J. Hepatol.* 2017; **67**: 145–72.
21 Blond E, Disse E, Cuerq C *et al*. EASL–EASD–EASO clinical practice guidelines for the management of non-alcoholic fatty liver disease in severely obese people: do they lead to over-referral? *Diabetologia* 2017; **60**: 1218–22.
22 Newsome PN, Cramb R, Davison SM *et al*. Guidelines on the management of abnormal liver blood tests. *Gut* 2018; **67**: 6–19.
23 Non-alcoholic fatty liver disease (NAFLD): assessment and management; NICE guidelines (NG49). In. National Institute for Health and Care Excellence: National Institute for Health and Care Excellence; 2016.
24 Xie Q, Zhou X, Huang P, Wei J, Wang W, Zheng S. The performance of enhanced liver fibrosis (ELF) test for the staging of liver fibrosis: a meta-analysis. *PLoS One* 2014; **9**: e92772.
25 Vali Y, Lee J, Boursier J *et al*. Enhanced liver fibrosis test for the non-invasive diagnosis of fibrosis in patients with NAFLD: a systematic review and meta-analysis. *J. Hepatol.* 2020; **73**: 252–62.
26 McInnes MDF, Moher D, Thombs BD *et al*. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA* 2018; **319**: 388–96.
27 Dvorak K, Stritesky J, Petrtyl J *et al*. Use of non-invasive parameters of non-alcoholic steatohepatitis and liver fibrosis in daily practice—an exploratory case-control study. *PLoS One* 2014; **9**: e111551.
28 Eddowes PJ, McDonald N, Davies N *et al*. Utility and cost evaluation of multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease. *Aliment. Pharmacol. Ther.* 2018; **47**: 631–44.
29 Nobili V, Parkes J, Bottazzo G *et al*. Performance of ELF serum markers in predicting fibrosis stage in pediatric non-alcoholic fatty liver disease. *Gastroenterology* 2009; **136**: 160–7.
30 Whiting PF, Rutjes AW, Westwood ME *et al*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann. Intern. Med.* 2011; **155**: 529–36.
31 Catanzaro R, Milazzo M, Arona S *et al*. Diagnostic accuracy of enhanced liver fibrosis test to assess liver fibrosis in patients with chronic hepatitis C. *Hepatobiliary Pancreat. Dis. Int.* 2013; **12**: 500–7.
32 Cobbold JF, Crossey MM, Colman P *et al*. Optimal combinations of ultrasound-based and serum markers of disease severity in patients with chronic hepatitis C. *J. Viral Hepat.* 2010; **17**: 537–45.
33 D'Ambrosio R, Degasperi E, Aghemo A *et al*. Serological tests do not predict residual fibrosis in hepatitis C cirrhotics with a sustained virological response to interferon. *PLoS One* 2016; **11**: e0155967.
34 Guechot J, Trocme C, Renversez JC, Sturm N, Zarski JP. Independent validation of the enhanced liver fibrosis (ELF) score in the ANRS HC EP 23 Fibrostar cohort of patients with chronic hepatitis C. *Clin. Chem. Lab. Med.* 2012; **50**: 693–9.
35 Fernandes FF, Ferraz ML, Andrade LE *et al*. Enhanced liver fibrosis panel as a predictor of liver fibrosis in chronic hepatitis C patients. *J. Clin. Gastroenterol.* 2015; **49**: 235–41.
36 Fujita K, Kuroda N, Morishita A *et al*. Fibrosis staging using direct serum biomarkers is influenced by hepatitis activity grading in hepatitis C virus infection. *J. Clin. Med.* 2018; **7**.
37 Martinez SM, Fernandez-Varo G, Gonzalez P *et al*. Assessment of liver fibrosis before and after antiviral therapy by different serum marker panels in patients with chronic hepatitis C. *Aliment. Pharmacol. Ther.* 2011; **33**: 138–48.
38 Parkes J, Guha IN, Roderick P *et al*. Enhanced liver fibrosis (ELF) test accurately identifies liver fibrosis in patients with chronic hepatitis C. *J. Viral Hepat.* 2011; **18**: 23–31.
39 Ragazzo TG, Paranagua-Vezozzo D, Lima FR *et al*. Accuracy of transient elastography-FibroScan®, acoustic radiation force impulse (ARFI) imaging, the enhanced liver fibrosis (ELF) test, APRI, and the FIB-4 index compared with liver biopsy in patients with chronic hepatitis C. *Clinics (Sao Paulo)* 2017; **72**: 516–25.
40 Tanwar S, Trembling PM, Hogan BJ *et al*. Biomarkers of hepatic fibrosis in chronic hepatitis C: a comparison of 10 biomarkers using 2 different assays for hyaluronic acid. *J. Clin. Gastroenterol.* 2017; **51**: 268–77.
41 Abdel-Hameed E, Rouster S, Kottilil S, Sherman K. The enhanced liver fibrosis (ELF)-Index predicts hepatic fibrosis superior to FIB4 and APRI in HIV/HCV infected patients. *Clin. Infect. Dis*. 2020; ciaa646. https://doi.org/10.1093/cid/ciaa646
42 Kim BK, Kim HS, Park JY *et al*. Prospective validation of ELF test in comparison with FibroScan and FibroTest to predict liver fibrosis in Asian subjects with chronic hepatitis B. *PLoS One* 2012; **7**: e41964.
43 Trembling PM, Lampertico P, Parkes J *et al*. Performance of enhanced liver fibrosis test and comparison with transient elastography in the identification of liver fibrosis in patients with chronic hepatitis B infection. *J. Viral Hepat.* 2014; **21**: 430–8.
44 Wong GL, Chan HL, Choi PC *et al*. Non-invasive algorithm of enhanced liver fibrosis and liver stiffness measurement with transient elastography for advanced liver fibrosis in chronic hepatitis B. *Aliment. Pharmacol. Ther.* 2014; **39**: 197–208.
45 Heo JY, Kim BK, Park JY *et al*. Combination of transient elastography and an enhanced liver fibrosis test to assess the degree of liver fibrosis in patients with chronic hepatitis B. *Gut Liver.* 2018; **12**: 190–200.
46 Anstee QM, Lawitz EJ, Alkhouri N *et al*. Noninvasive tests accurately identify advanced fibrosis due to NASH: baseline data from the STELLAR trials. *Hepatology* 2019; **70**: 1521–30.
47 Guillaume M, Moal V, Delabaudiere C *et al*. Direct comparison of the specialised blood fibrosis tests FibroMeter V2G and enhanced liver fibrosis score in patients with non-alcoholic fatty liver disease from tertiary care centres. *Aliment. Pharmacol. Ther.* 2019; **50**: 1214–22.
48 Inadomi C, Takahashi H, Ogawa Y *et al*. Accuracy of the enhanced liver fibrosis test, and combination of the enhanced liver fibrosis and non-invasive tests for the diagnosis of advanced liver fibrosis in patients with non-alcoholic fatty liver disease. *Hepatol. Res.* 2020; **50**: 682–92.
49 Lykiardopoulos B, Hagström H, Fredrikson M *et al*. Development of serum marker models to increase diagnostic accuracy of advanced fibrosis in nonalcoholic fatty liver disease: the new LINKI algorithm compared with established algorithms. *PLoS One* 2016; **11**: e0167776.

50 Miele L, De Michele T, Marrone G *et al*. Enhanced liver fibrosis test as a reliable tool for assessing fibrosis in nonalcoholic fatty liver disease in a clinical setting. *Int. J. Biol. Markers* 2017; **32**: e397–402.
51 Staufer K, Halilbasic E, Spindelboeck W *et al*. Evaluation and comparison of six noninvasive tests for prediction of significant or advanced fibrosis in nonalcoholic fatty liver disease. *United Eur. Gastroenterol. J.* 2019; **7**: 1113–23.
52 Thiele M, Madsen BS, Hansen JF, Detlefsen S, Antonsen S, Krag A. Accuracy of the enhanced liver fibrosis test *vs* fibrotest, elastography, and indirect markers in detection of advanced fibrosis in patients with alcoholic liver disease. *Gastroenterology* 2018; **154**: 1369–79.
53 Madsen BS, Thiele M, Detlefsen S *et al*. Prediction of liver fibrosis severity in alcoholic liver disease by human microfibrillar-associated protein 4. *Liver Int.* 2020; **40**: 1701–12.
54 Agrawal S, Hoad CL, Francis ST, Guha IN, Kaye P, Aithal GP. Visual morphometry and three non-invasive markers in the evaluation of liver fibrosis in chronic liver disease. *Scand. J. Gastroenterol.* 2017; **52**: 107–15.
55 Crespo G, Fernandez-Varo G, Marino Z *et al*. ARFI, FibroScan, ELF, and their combinations in the assessment of liver fibrosis: a prospective study. *J. Hepatol.* 2012; **57**: 281–7.
56 Dyvorne HA, Jajamovich GH, Bane O *et al*. Prospective comparison of magnetic resonance imaging to transient elastography and serum markers for liver fibrosis detection. *Liver Int.* 2016; **36**: 659–66.
57 Fagan KJ, Pretorius CJ, Horsfall LU *et al*. ELF score ≥9.8 indicates advanced hepatic fibrosis and is influenced by age, steatosis and histological activity. *Liver Int.* 2015; **35**: 1673–81.
58 Friedrich-Rust M, Rosenberg W, Parkes J, Herrmann E, Zeuzem S, Sarrazin C. Comparison of ELF, FibroTest and FibroScan for the non-invasive assessment of liver fibrosis. *BMC Gastroenterol.* 2010; **10**: 103.
59 Irvine KM, Wockner LF, Hoffmann I *et al*. Multiplex serum protein analysis identifies novel biomarkers of advanced fibrosis in patients with chronic liver disease with the potential to improve diagnostic accuracy of established biomarkers. *PLoS One* 2016; **11**: e0167001.
60 Lee MH, Cheong JY, Um SH *et al*. Comparison of surrogate serum markers and transient elastography (FibroScan) for assessing cirrhosis in patients with chronic viral hepatitis. *Dig. Dis. Sci.* 2010; **55**: 3552–60.
61 Stasi C, Tsochatzis EA, Hall A *et al*. Comparison and correlation of fibrosis stage assessment by collagen proportionate area (CPA) and the ELF panel in patients with chronic liver disease. *Dig. Liver Dis.* 2019; **51**: 1001–7.
62 Wahl K, Rosenberg W, Vaske B *et al*. Biopsy-controlled liver fibrosis staging using the enhanced liver fibrosis (ELF) score compared to transient elastography. *PLoS One* 2012; **7**: e51906.
63 Abozeid M, Alsebaey A, Abdelsameea E *et al*. High efficacy of generic and brand direct acting antivirals in treatment of chronic hepatitis C. *Int. J. Infect. Dis.* 2018; **75**: 109–14.
64 Glen J, Floros L, Day C, Pryke R. Non-alcoholic fatty liver disease (NAFLD): summary of NICE guidance. *BMJ* 2016; **354**: i4428.
65 Tanwar S, Srivastava A, Rosenberg W. Errors in modeling misrepresent the utility of the enhanced liver fibrosis test in the management of non-alcoholic fatty liver disease. *J. Hepatol.* 2020; **S0168–8278**: 30462–1.
66 de Vries E, Färkkilä M, Milkiewicz P *et al*. Enhanced liver fibrosis test predicts transplant-free survival in primary sclerosing cholangitis, a multi-centre study. *Liver Int.* 2017; **37**: 1554–61.
67 Wang J, Li J, Zhou Q *et al*. Liver stiffness measurement predicted liver-related events and all-cause mortality: a systematic review and nonlinear dose-response meta-analysis. *Hepatol Commun.* 2018; **2**: 467–76.
68 Salomone F, Micek A, Godos J. Simple scores of fibrosis and mortality in patients with NAFLD: a systematic review with meta-analysis. *J. Clin. Med.* 2018; **7**: 219.
69 Houot M, Ngo Y, Munteanu M, Marque S, Poynard T. Systematic review with meta-analysis: direct comparisons of biomarkers for the diagnosis of fibrosis in chronic hepatitis C and B. *Aliment. Pharmacol. Ther.* 2016; **43**: 16–29.
70 Li Y, Huang Y, Wang Z *et al*. Systematic review with meta-analysis: the diagnostic accuracy of transient elastography for the staging of liver fibrosis in patients with chronic hepatitis B. *Aliment. Pharmacol. Ther.* 2016; **43**: 458–69.
71 Pavlov C, Casazza G, Nikolova D, Tsochatzis E, Gluud C. Systematic review with meta-analysis: diagnostic accuracy of transient elastography for staging of fibrosis in people with alcoholic liver disease. *Aliment. Pharmacol. Ther.* 2016; **43**: 575–85.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1: Data extraction instrument.
Table S2: Different grading systems for liver fibrosis assessed by biopsy compared to ELF score.
Table S3: – sensitivities and specificities with confidence intervals for each paper at presented thresholds, as calculated by contingency table data.
Table S4: QUADAS 2 tool28 guide.