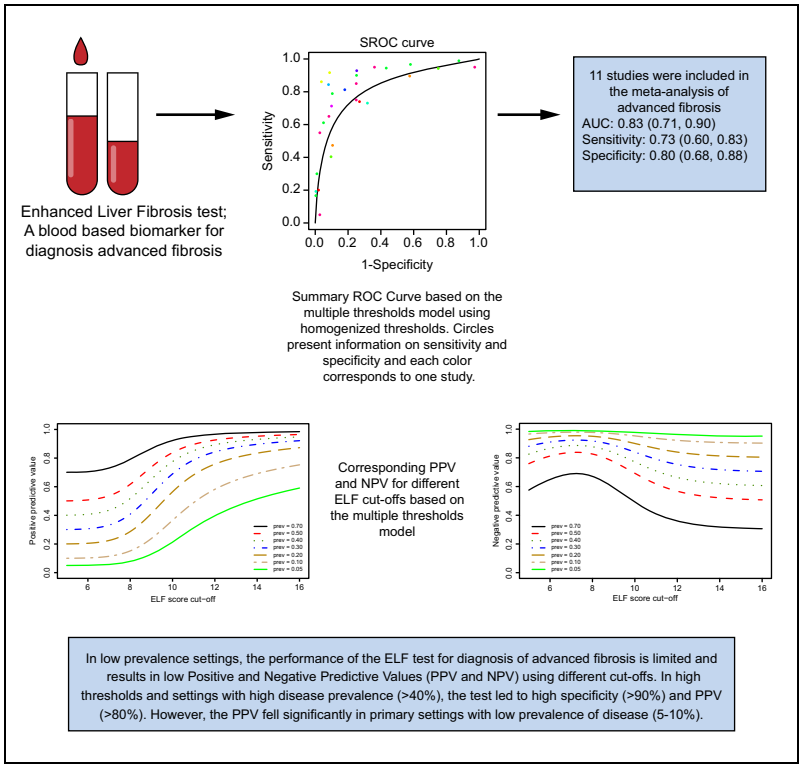
Research Article
NAFLD and Alcohol-Related Liver Diseases

JOURNAL
OF HEPATOLOGY

Enhanced liver fibrosis test for the non-invasive diagnosis of fibrosis in patients with NAFLD: A systematic review and meta-analysis

Graphical abstract



Highlights

- The ELF test is measured using various algorithms that combine the same components.
- Different predefined cut-offs are suggested in the guidelines (and by the manufacturer) for the ELF test.
- ELF has a high sensitivity but limited specificity to exclude NAFLD-related fibrosis.
- ELF had limited performance for diagnosing fibrosis, especially in low-prevalence settings.

Authors

Yasaman Vali, Jenny Lee, Jérôme Boursier, …, Quentin M. Anstee, Patrick M. Bossuyt, Mohammad Hadi Zafarmand

Correspondence

y.vali@amsterdamumc.nl (Y. Vali).

Lay summary

The enhanced liver fibrosis test has been suggested as a non-invasive blood test to aid the diagnosis of severe liver fibrosis in patients with non-alcoholic fatty liver disease (NAFLD). Our study results showed that the test has a high negative predictive value, especially in populations with low disease prevalence (likely encountered in primary care); so, it can exclude advanced fibrosis in patients with NAFLD. However, when prevalence is low, the positive predictive value of the enhanced liver fibrosis test is low, suggesting that additional strategies may be needed to make a positive diagnosis in such settings.

https://doi.org/10.1016/j.jhep.2020.03.036
 J. Hepatol. 2020, 73, 252–262

Research Article
NAFLD and Alcohol-Related Liver Diseases

JOURNAL OF HEPATOLOGY

Enhanced liver fibrosis test for the non-invasive diagnosis of fibrosis in patients with NAFLD: A systematic review and meta-analysis

Yasaman Vali[1,*], Jenny Lee[1], Jérôme Boursier[2,3], René Spijker[4,5], Jürgen Löffler[6], Joanne Verheij[7], M. Julia Brosnan[8], Zsolt Böcskei[9], Quentin M. Anstee[10,11], Patrick M. Bossuyt[1], Mohammad Hadi Zafarmand[1], and the LITMUS systematic review team[†]

[1]*Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam UMC, University of Amsterdam, The Netherlands;* [2]*Hepato-Gastroenterology Department, Angers University Hospital, Angers, France;* [3]*HIFIH Laboratory, UPRES EA3859, Angers University, Angers, France;* [4]*Medical Library AMC, Amsterdam, the Netherlands;* [5]*Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands;* [6]*Novartis Pharma AG, Basel, Switzerland;* [7]*Department of Pathology, Academic Medical Center, Amsterdam, The Netherlands;* [8]*Internal Medicine Research Unit, Pfizer Inc, Cambridge, MA, USA;* [9]*Sanofi R&D, Translational Sciences Unit, Chilly Mazarin, France;* [10]*The Newcastle Liver Research Group, Institute of Cellular Medicine, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK;* [11]*Newcastle NIHR Biomedical Research Centre, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK*

Background & Aims: The enhanced liver fibrosis (ELF) test has been proposed for the non-invasive assessment of advanced fibrosis in patients with non-alcoholic fatty liver disease (NAFLD). We performed a systematic review to estimate the accuracy of this test against biopsy.

Methods: In this systematic review, we searched MEDLINE, Embase, Web of Science and the Cochrane Library for studies that included patients with NAFLD and that used both liver biopsy (as the reference standard) and the ELF test. Two authors independently screened the references, extracted the data and assessed the quality of included studies. Due to the variation in reported thresholds, we used a multiple thresholds random effects model for meta-analysis (diagmeta R-package).

Results: The meta-analysis of 11 studies reporting advanced fibrosis and 5 studies reporting significant fibrosis showed that the ELF test had a sensitivity of >0.90 for excluding fibrosis at a threshold of 7.7. However, as a diagnostic test at high thresholds, the test only achieved specificity and positive predictive value >0.80 in very high prevalence settings (>50%). To achieve a specificity of 0.90 for advanced and significant fibrosis, thresholds of 10.18 (sensitivity: 0.57) and 9.86 (sensitivity: 0.55) were required, respectively.

Conclusion: The ELF test showed high sensitivity but limited specificity to exclude advanced and significant fibrosis at low cut-offs. The diagnostic performance of the test at higher thresholds was found to be more limited in low-prevalence settings. We conclude that clinicians should carefully consider the likely disease prevalence in their practice setting and adopt suitable test thresholds to achieve the desired performance.

Lay summary: The enhanced liver fibrosis test has been suggested as a non-invasive blood test to aid the diagnosis of severe liver fibrosis in patients with non-alcoholic fatty liver disease (NAFLD). Our study results showed that the test has a high negative predictive value, especially in populations with low disease prevalence (likely encountered in primary care); so, it can exclude advanced fibrosis in patients with NAFLD. However, when prevalence is low, the positive predictive value of the enhanced liver fibrosis test is low, suggesting that additional strategies may be needed to make a positive diagnosis in such settings.

Introduction

Non-alcoholic fatty liver disease (NAFLD) is a potentially progressive disorder associated with the clinical features of metabolic syndrome.[1,2] With a global prevalence of 25%, it is now the leading cause of chronic liver disease worldwide.[3,4] NAFLD represents a wide spectrum of disease and there are a large number of patients who can develop progressive liver fibrosis and non-alcoholic steatohepatitis (NASH).[5,6] Evidence shows that any advance in fibrosis stage can exponentially increase liver-related mortality.[6,7] As the progression of liver fibrosis is considered the most important predictor of NAFLD-related outcomes, early identification of patients with NAFLD and advanced fibrosis (F3/4) is recommended by international guidelines[8–10] and is a key area of interest for clinical trial recruitment.[11]

The current reference standard for diagnosis of NASH and hepatic fibrosis is liver biopsy. However, it is invasive, resource intensive and prone to sampling error if not of adequate size (16 Gauge needle biopsy, 20 mm length).[12–14] Moreover, it carries a small but significant risk of complications,[12,14] which makes it less suitable for diagnosis in clinical practice or in drug development settings.

In recent years, attention has been given to non-invasive NAFLD biomarkers. Several biomarkers have been developed

Keywords: Non-alcoholic fatty liver disease; Fibrosis; Non-alcoholic steatohepatitis; Enhanced liver fibrosis test; Biomarker; Meta-analysis.
Received 16 October 2019; received in revised form 18 March 2020; accepted 20 March 2020; available online 8 April 2020
* Corresponding author. Address: Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Room J1b-227, Amsterdam UMC, Location AMC, Meibergdreef 9, 1105AZ Amsterdam, The Netherlands. Tel.: +31(0)20 5668520.
E-mail address: y.vali@amsterdamumc.nl (Y. Vali).
https://doi.org/10.1016/j.jhep.2020.03.036

and studied, ranging from simple blood-based biomarkers to more complex panels including imaging modalities. The recent National Institute of health and Care Excellence (NICE) guideline on NAFLD suggests the use of the enhanced liver fibrosis (ELF) test, a non-invasive blood-derived biomarker, to aid the diagnosis of advanced fibrosis in the patients with NAFLD.[15]

The ELF test is a panel of markers that consists of 3 components: type III procollagen peptide (PIIINP), hyaluronic acid (HA), and tissue inhibitor of metalloproteinase-1 (TIMP1).[16,17] It is notable that, whilst the test has retained the same name, the formula and components included have been altered sequentially so not all studies reflect the performance of the currently available test. Initially, the combination of these markers with age was introduced as Original European Liver Fibrosis test in 2004.[16] Later, the test was simplified by removing age and a new algorithm for generating the ELF test was published in 2008.[17] This algorithm was subsequently revised again by the Siemens company and reported based on 2 different Siemens platforms.[18]

The clinical performance of this biomarker in NAFLD has been reported by several studies, with variable results. The health-economic model to support the development of the NICE guideline reported the highest diagnostic accuracy of the ELF test compared to 13 other diagnostic strategies. The NICE guidelines consequently recommend that "the ELF test should be considered in people who have been diagnosed with NAFLD to test for advanced fibrosis", suggesting an ELF score of 10.51 as the cut-off value.[19] The estimate for the health-economic model was proposed based on a single study, in a tertiary, pediatric setting.[20] Most other studies used manufacturer recommended thresholds for ruling advanced F3/4 fibrosis out (7.7) or in (9.8) respectively, and reported different levels of performance of this test.[17,21–35]

The variability in reported results brings into question all recommendations regarding the ELF test as a diagnostic tool for NAFLD-related fibrosis, with consequences for clinical diagnoses, treatment decisions, and drug development. We aimed to perform a systematic review and meta-analysis of the published studies, to provide summary estimates of the accuracy of the ELF test for diagnosing liver fibrosis and NASH in patients with NAFLD, at the recommended thresholds, based on the currently available evidence. In addition, we sought to evaluate alternative cut-off values.

Materials and methods

The present study was conducted as part of a larger multi-center project named LITMUS (Liver Investigation: Testing Marker Utility in Steatohepatitis), which is funded by the European Union IMI2 scheme aiming to develop, validate and qualify a defined set of biomarkers that enable detection of NASH and fibrosis.

Inclusion criteria

Types of studies and participants

Studies, reported in peer-reviewed journals or conference abstracts, that included patients (≥18 years) with biopsy-proven or suspected NAFLD, with paired liver histology and index biomarker data were potentially eligible for this review. We made no further restrictions based on either year or language. Studies of participants with mixed etiologies were only included if the performance of the biomarker was separately reported in patients with NAFLD.

Target condition

The target conditions were NASH (with or without fibrosis) and staging of fibrosis (independent of NASH activity). We intended to capture the stages of liver fibrosis using the F0 to F4 scale, as defined by the NASH Clinical Research Network (CRN) staging system or other scoring systems.[36] Table S1 and S2 show the different fibrosis scoring systems[36–43] and various histological scoring systems for characterizing NAFLD progression.[36,44–47]

Index biomarker

Using the same components, different algorithms have been developed for ELF test, by Guha *et al.* 2008[17] and later by Siemens (for their 2 different platforms[18]) to produce a score to assess liver fibrosis:

1. Guha algorithm:

- -7.412 + 0.681 ln(HA) + 0.775 ln(PIIINP) + 0.494 ln(TIMP1)

2. Siemens algorithms:

- *Using the ADVIA Centaur and ADVIA Centaur XP/XPT Systems as an advanced automated immunoassay analyzer:*
  2.278 + 0.851 ln(HA) + 0.751 ln(PIIINP) + 0.394 ln(TIMP1)
- *Using the ADVIA Centaur CP System as an immunoassay test instrument for mid-volume labs:*
  2.494 + 0.846 ln(HA) + 0.735 ln(PIIINP) + 0.391 ln(TIMP1)

These different algorithms produce highly correlated results. The algorithms showed a very high positive correlation (R^2 = 0.995 and 0.993, respectively; Fig. S1) when compared between 2 independent studies (the first among 502 patients with biopsy-proven NAFLD from 2 university hospitals in France; the second among 532 patients with NAFLD from different counties), for which we had access to the data. We therefore used the following regression equation to harmonize the test results and to convert all thresholds to those of the Siemens algorithm:

Siemens results = (Guha results + 8.6498)/0.8854

Reference standard

We only included studies in which histological assessment of liver biopsy was used as the reference standard for grading and staging of NAFLD.

Exclusion criteria

Studies were excluded if they included participants with coexisting liver disease (*e.g.* viral hepatitis), decompensated cirrhosis, or addressed a different context of use. Letter/commentary designs were also excluded. In addition, we excluded publications that did not provide enough data to calculate diagnostic accuracy estimates.

Search methods

The following databases were searched:

- MEDLINE via OVID
- PubMed
- Embase via OVID (including conference abstracts)
- Science Citation Index
- CENTRAL (The Cochrane Library)

A sensitive search strategy was developed in close collaboration with our search specialist (RS). Table S3 shows the search strategy for MEDLINE. The final search strategy was adapted for different databases and initially used in August 2018. It was

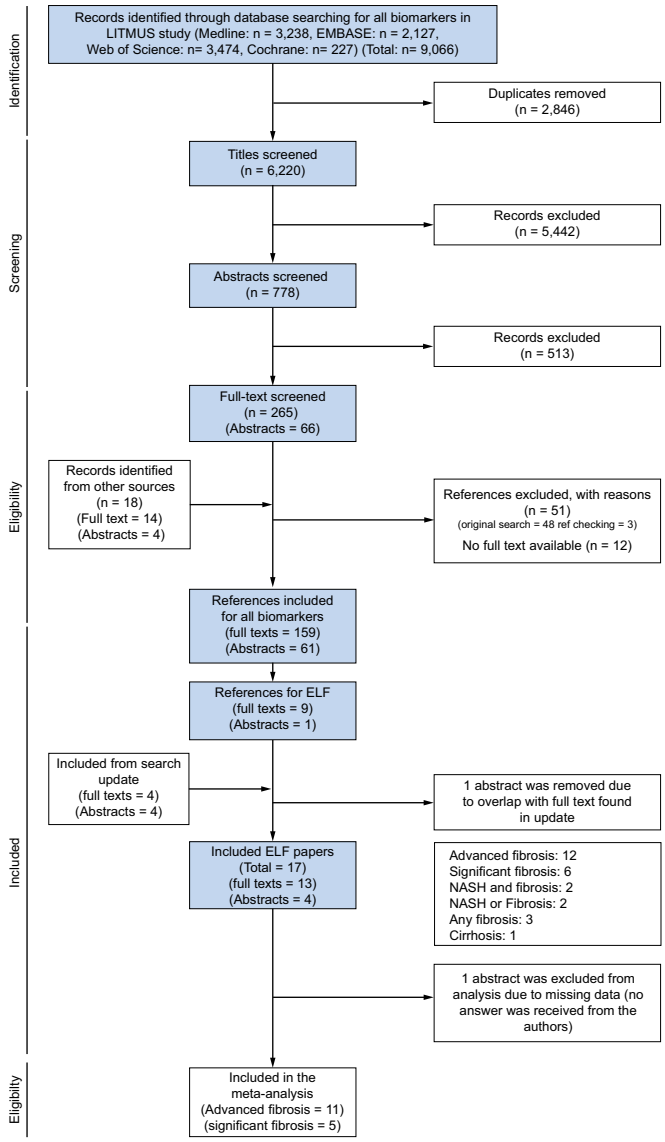


Fig 1. PRISMA flow diagram of primary studies.

updated specifically for the ELF test in February 2019 and December 2019 (Fig. 1).

To identify additional studies, we screened the reference list of related systematic reviews and the included studies. In addition, we contacted the academic and industry partners within the LITMUS consortium for any studies that were missed by our search strategy.

Data collection and analysis

Selection of studies and data extraction

Using Endnote, duplicate records were removed and titles of all records were screened by 1 reviewer (YV) while the second reviewer screened 10% of the titles independently. Abstracts and full texts were sought and independently assessed by 2 authors (YV and JL). In case of any disagreement, consensus was reached by discussion, first between 2 reviewers and if necessary, with a senior member of the team (MHZ). The screening phase was managed using the Rayyan software (https://rayyan.qcri.org).

The following data were extracted by 1 author (YV/JL) and cross checked by the other author (JL/YV): study group characteristics, index test and reference test features, number of true and false positives, and true and false negatives for constructing classification tables.

Assessment of methodological quality

Only studies with full-text reports were assessed for methodological quality. Two authors (YV and JL) independently evaluated the quality of included studies using the QUADAS-2 tool.[48]

Statistical analysis

Classification tables were extracted or reconstructed for the performance of the index biomarker of each predefined target condition. For the analyses, we extracted the accuracy data on each cut-off point for which the data was available or could be calculated.

Different studies could contribute a varying number of thresholds, as well as different sets of thresholds. Estimates of sensitivity, specificity, and corresponding 95% CIs were generated and graphically illustrated in forest plots.

We planned to conduct a meta-analysis whenever more than 3 studies with enough information to create classification tables were available. We used a linear mixed effects model for modeling the multiple thresholds data of the individual studies, as recently proposed in “diagmeta” package in R. The multiple thresholds model is a multilevel random effects model that enables the calculation of summarized sensitivities and specificities of different cut-off points, and the calculation of the predictive values, given the prevalence of the target condition of interest.[49,50]

Sensitivity and specificity were combined at every recommended threshold and a multiple thresholds summary receiver-operating characteristic (SROC) curve was produced. Additionally, we obtained positive and negative predictive values (PPVs/NPVs). We also calculated thresholds of the ELF test required to achieve pre-specified (high) values of sensitivity and specificity. We defined minimally acceptable performance levels as 0.8 for both sensitivity and specificity, for ELF to exceed the performance of other NAFLD-related fibrosis screening and diagnostic biomarkers.[51]

The 95% CIs for sensitivity and specificity were estimated using the delta method.[50] We used a log-logistic model in our analyses to provide estimates of the 2 cumulative distribution functions of the test results, 1 within the disease-free and 1 within the diseased individuals, across all studies, accounting for the between study heterogeneity and correlation between the groups. Each data point was weighted with the inverse variance of the respective logit-transformed proportion.

R for Windows (Version 3.6.0; R Foundation for Statistical Computing, Vienna, Austria) was used in all analyses. For sensitivity analysis, we investigated the influence of disease severity among study groups, by removing a study with a very high prevalence of advanced fibrosis.[52] We also assessed the effect of the test-biopsy time interval by removing studies with a long time interval between test and liver biopsy (mean: >12 months).[26,34] Heterogeneity was assessed based on visual assessment of forest plots and ROC curves.

We did not attempt to construct funnel plots as is well known for systematic reviews of test accuracy studies, statistical tests based on funnel plot asymmetry cannot discriminate between

publication bias and other sources of asymmetry, like the effect of including multiple thresholds.[53]

The protocol of the full systematic review is available in PROSPERO: CRD42018106821. This study was reported using the PRISMA-DTA statement[54] (see Table S4).

Results

Search results

The initial search of the electronic databases resulted in 9,066 references. After removing duplicates, we screened 6,220 titles and 778 abstracts. We found 265 full-text reports from the electronic searches and 18 studies from other sources. In total, we were able to include 17 studies that had evaluated the accuracy of the ELF test in patients with NAFLD: 13 with full-text reports[17,23–30,34,35,52,55] and 4 only reported in abstracts[22,31–33] (Fig. 1). Reasons for exclusion are provided in Table S5.

Study characteristics

Target population

Major characteristics of the included studies are summarized in Table 1. The mean or median age of the participants included in these studies ranged from 42 to 60 years. Four studies included patients with suspected NAFLD,[17,29,30,55] 2 studies included morbidly obese patients,[27,35] while the remaining 11 studies evaluated patients with biopsy-proven NAFLD. There was a noticeable heterogeneity in severity of the disease among the included patients in the individual studies. The prevalence of advanced fibrosis ranged from 18% to 71%. The highest prevalence of advanced fibrosis was observed in a study that had targeted enrollment of patients in the context of therapeutic trials targeting bridging fibrosis (F3) or compensated cirrhosis (F4) with concomitant NASH as part of 2 phase III therapeutic clinical trials.[52]

Target conditions

The number of studies for each target condition is shown in Table S7. Accuracy in detecting advanced fibrosis (F ≥3) was reported by 11 studies[17,22–24,28–30,33,34,52,55]; in detecting significant fibrosis (F ≥2) by 5[17,24,26,35,55] studies. Only 1 study reported the performance of the test in detecting cirrhosis.[52] Almost all studies used NASH CRN (except 1 study that reported Brunt criteria). The studies reporting on advanced fibrosis had recruited 4,452 patients, of which 2,655 patients had fibrosis F ≥3. The studies reporting on significant fibrosis had included 550 patients with NAFLD, among whom 203 had fibrosis F ≥2. Tables S8 and S9 provide baseline characteristics of the 2 study groups included in our meta-analyses.

ELF test algorithms and thresholds

Although the majority of the studies used the Siemens formula, we found 3 studies that reported thresholds based on the Guha formula.[17,23,35] The thresholds reported by these studies were converted to Siemens scale. Studies reported on different thresholds, with a total of 8 studies reporting test accuracy at more than 1 threshold: 7 studies included in the meta-analysis of advanced fibrosis[17,21–24,28,34] and 4 studies in significant fibrosis.[17,24,35,55]

Biopsy characteristics

Not all studies provided detailed information about the biopsy. Only 6 studies reported the length of the biopsy samples[22,24,25,28,29,55]; 3 studies reported the needle gauge.[24,25,28] Biopsy samples were evaluated by a single pathologist in most of the studies; 5 studies reported evaluations by 2 or more pathologists and only 4 studies used hepatopathologists. The time interval between biopsy and the blood sampling varied significantly. Two studies reported long test-biopsy time intervals. In 1 study the time interval between liver biopsy and participant inclusion, when the blood samples were collected, was up to 48 months.[26] In the other study, patients with NAFLD had undergone biopsy 1.62 ± 1.75 years (mean ± SD) before blood was sampled.[34] See Table S6 for details.

Methodological quality assessment

The methodological quality assessment results are summarized in Fig. S2 and further illustrated for individual studies in Fig. S3. Only 1 study had a low risk of bias in all 4 domains.[28] Four studies were scored high for risk of bias on the patient selection domain, 2 on index test, and 3 on flow and timing. None of the studies was scored as high risk of bias for the reference standard domain. One study was considered to have concerns about applicability, both in terms of patient selection and index test,[27] since it had recruited from obese patients (14 males and 43 females), who underwent bariatric surgery in 1 hospital. The study proposed only 1 cut-off value for diagnosing NASH with any levels of fibrosis in obese patients with NAFLD.

As the number of studies included in the series of meta-analyses was limited (with a maximum of 10 studies for advanced fibrosis) sources of clinical heterogeneity were not further explored.

Overall accuracy of ELF test for advanced fibrosis (≥F3)

Fig. S4 shows forest plots for the diagnostic accuracy of ELF test in detecting advanced fibrosis. The studies were not consistent in reporting the low and high thresholds. The forest plots illustrate the heterogeneity in thresholds reported by each study and their corresponding sensitivities and specificities.

ELF test performance in published recommended thresholds

The application of the multiple thresholds model leads to the SROC curve (Fig. 2), which enables the estimation of sensitivities and specificities at different thresholds, for instance for the predefined thresholds: 7.7 and 9.8 (low and high thresholds recommended by Siemens) and 10.51 (the high threshold recommended by NICE guideline). The area under the SROC curve (AUC) for detecting advanced fibrosis is 0.83 (95% CI 0.71–0.90). Table 2 shows the accuracy of the ELF test for diagnosing advanced fibrosis at the proposed thresholds. The recommended lower threshold of 7.7 showed a high sensitivity of 0.93 (95% CI 0.82–0.98) with a specificity of 0.34 (95% CI 0.13–0.65).

For the high thresholds, we observed a specificity of 0.86 (95% CI 0.77–0.92) and sensitivity of 0.65 (95% CI 0.49–0.77) at ELF >9.8 and a specificity of 0.93 (95% CI 0.85–0.96) and sensitivity of 0.51 (95% CI 0.31–0.70) at the NICE recommended threshold of ELF >10.51.

Modeling ELF test performance for different clinical settings and disease prevalence

In clinical practice, there is a need to understand the probability that a NAFLD patient has or does not have advanced fibrosis when an ELF test result exceeds a certain threshold. To address this, we additionally used the multiple thresholds model and

Table 1. Studies included in the systematic review and meta-analysis.

Study	Setting	Population	N	M/F	Mean Age	Mean BMI	ELF formula		Target condition/N (%)
Guha 2008	Outpatients Tertiary	Suspected NAFLD	192	123/69	48.7	32.4	Guha	−7.412 + (ln(HA)*0.681) + (ln(PIIINP)*0.775) + (ln(TIMP1)*0.494)	Advanced fibrosis (F3-4)/44 (22.9) Significant fibrosis (F ≥2)/77 (40.1) Any fibrosis/113 (58.9)
Younossi 2011	Outpatients Hospital	Obese Biopsy-proven NAFLD	79	18/61	42.3	47.6	Guha	NR	Significant fibrosis (F ≥2)/16 (20.3) Any fibrosis/39 (49.4)
Dvorak 2014	Outpatients Hospital	Biopsy-proven NAFLD	56	NR	45.5	30.5	Guha	−7.412 + (ln(HA)*0.681) + (ln(PIIINP)* 0.775) + (ln(TIMP1)*0.494)	Advanced fibrosis (F3-4)/17 (30.3)
Karlas 2015	Outpatients Hospital	Non-bariatric NAFLD	48	24/24	55.3	27.5	Siemens	2.494 + ln(CHA) + ln(CPIIINP) + ln(CTIMP−1)	Significant fibrosis (F ≥2)/8 (16.6)
Lykiardopoulos 2016	Outpatients Hospital	Biopsy-proven NAFLD	158	117/41	60*	28.7*	Siemens	2.2781 + 0.851 × ln(HA) +10.751 × l (PIIINP) +10.934 × ln(TIMP 1)	Advanced fibrosis (F3-4)/38 (24)
Sanyal 2016	NR	Biopsy-proven NAFLD	216	96/120	52.1	31.2	Siemens	NR	NASH with fibrosis (NAS ≥4; F2-3)/95 (44)
Lopez 2017	Inpatients Hospital	Bariatric	57	14/43	44.0	49.1	Siemens	2.278 + 0.851 ln(HA) + 0.751 ln(PIIINP) + 0.394 ln(TIMP-1)]	NASH or any Fibrosis/29 (50.9)
Miele 2017	Outpatients Hospital	Suspected NAFLD	82	62/20	46	28.0*	Siemens	2.494 + 0.846 ln(CHA) + 0.735 ln(CPIIINP) + 0.391 ln(CTIMP-1)	Advanced fibrosis (F3-4)/15 (18.3)
Boursier 2018	NR	Biopsy-proven NAFLD	417	247/170	56.1	33.3	Siemens	2.278 + 0.851 ln(HA) + 0.751 ln(PIIINP) + 0.394 ln(TIMP1)	Advanced fibrosis (F3-4)/167 (40)
Eddowes 2018	Out and Inpatients Hospital	Biopsy-proven NAFLD	54	28/26	54*	33.6	Siemens	NR	NASH or any Fibrosis/38 (76) Advanced fibrosis (F3-4)/25 (50) Significant fibrosis (F ≥2)/34 (68)
Itoh 2018	Outpatients Hospital	Biopsy-proven NAFLD	400	195/205	56*	27.3*	Siemens	2.278 + 0.851 ln(HA) + 0.751 ln(PIIINP) + 0.394 ln(TIMP-1)]	Any fibrosis (F ≥1)/334 (83.5)
Shulze 2018	NR	Biopsy-proven NAFLD	74	NR	NR	NR	NR	NR	Significant fibrosis (F ≥2)/13 (17.6)
Staufer 2018	Outpatients Hospital	Biopsy-proven NAFLD	122	NR	NR	NR	Siemens	2.494 + 0.846 ln(HA) + 0.735 ln(PIIINP) + 0.391 ln(TIMP1)	Advanced fibrosis (F3-4)/34 (27.9)
Welsh 2018	Inpatients Hospital	Biopsy-proven NAFLD	26	13/13	50	34	Siemens	NR	Advanced fibrosis (F3-4)/9 (34.6)
Polyzos 2019	Outpatients Hospital	Suspected NAFLD	31	9/22	53.7	33	Siemens	2.278 + 0.851 × ln(HA) + 0.751 × ln(PIIINP) + 0.394 × ln(TIMP-1)	Advanced fibrosis (F3-4)/7 (22.6)
Anstee 2019	Outpatients Hospital	Biopsy-proven NASH	3173	2010/2457	58*	NR	Siemens	2.278 + 0.851 ln(HA) + 0.751 ln(PIIINP) + 0.394 ln(TIMP1)	Advanced fibrosis (F3-4)/2,249 (70.9) Cirrhosis (F4)/1,274 (40.2)
Staufer 2019	Outpatients Hospital	Suspected NAFLD	181	101/80	52*	30.5*	Siemens	2.278 + 0.851 ln(HA) + 0.751 ln(PIIINP) + 0.394 ln(TIMP1)	Advanced fibrosis (F3-4)/46 (25.4) Significant fibrosis (F ≥2)/68 (37.6) NASH with advanced fibrosis (NAS ≥4; F = 3-4)/32 (17.7)

*Median. BMI, body mass index; HA, hyaluronic acid; NAFLD, non-alcoholic fatty liver disease; NAS, NAFLD activity score; NASH, non-alcoholic steatohepatitis; NR, not reported; PIIINP, type III procollagen peptide; TIMP, tissue inhibitor of metalloproteinase-1.

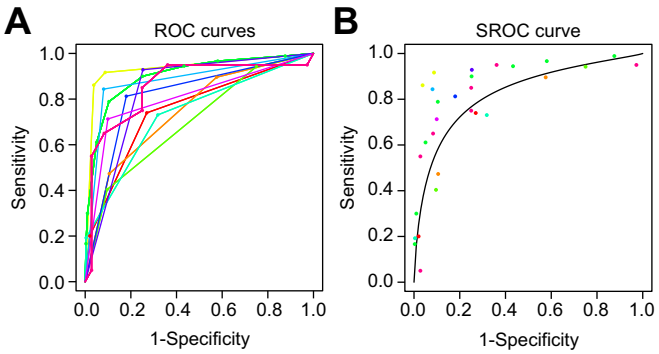


Fig. 2. Test performance for detecting advanced fibrosis. (A) Multiple threshold ROC curves and (B) Multiple threshold SROC curve based on the multiple thresholds model using homogenized thresholds. Circles represent information on sensitivity and specificity. AUC: 0.83 (0.71, 0.90). Max Y-index results: cut-off: 9.37, sensitivity: 0.73 (0.60, 0.83), specificity: 0.80 (0.68, 0.88). AUC, area under the ROC/SROC curve; ROC, receiver-operating characteristic; SROC, summary receiver-operating characteristic.

calculated the PPVs and NPVs related to the full range of threshold values for different levels of prevalence (Fig. 2). Employing an ELF threshold of 7.7, the highest NPV of 0.99 was observed when using the test in a low-prevalence setting, *i.e.* when the prevalence of advanced fibrosis was no more than 5% (Table 2, Fig. 3B). We additionally reported the performance of the ELF test for ruling out advanced fibrosis at prevalence levels equivalent to those that may be encountered in primary and secondary/tertiary care settings, with prevalence values ranging between 5% to 50%, leading to NPVs between 0.99 and 0.83, respectively.

Table 2 and Fig. 3A show the performance of the recommended high-cut-off thresholds. The high threshold of 9.8 had a mean specificity of 0.86 (95% CI 0.77–0.92), and PPV of 0.82 and 0.91 at disease prevalences of 50% and 70%, respectively. Adopting the higher threshold of 10.51, specificity was 0.93 (95% CI 0.85–0.96) with a PPV of 0.82 when the disease prevalence was 40%, but the PPV fell to just 0.26 at the 5% prevalence level, more likely to be encountered in primary care settings. An even higher threshold of the ELF test (11.3), showed a specificity of 0.96 (95% CI 0.90–0.99), and PPV would be 0.81 in case of a disease prevalence of at least 30%, while this fell to 0.34 at the 5% prevalence level.

Desired thresholds for diagnosis of advanced fibrosis

Table 3 shows the desired thresholds of the ELF test for different fixed high sensitivities (Table 3A) and specificities (Table 3B). These results were consistent with the findings of the previous analysis. In low cut-offs, when the prevalence was less than 50%, the ELF test showed high sensitivities (0.90), resulting in high NPVs, ranging from 0.82 to 0.99. However, none of the new high thresholds showed a high PPV (>0.80) with fixed specificities (0.90, 0.95 and 0.98) for disease prevalences of 5–20%. At the highest threshold of 12.01, at a fixed specificity of 0.98, the ELF test had a PPV >0.80 in settings with a prevalence of at least 30%.

Overall accuracy of ELF test for significant fibrosis (≥F2)

Five studies were included in the meta-analysis of significant fibrosis (Fig. S5). Fig. 4 shows the SROC curve for the significant fibrosis meta-analysis, with an AUC of 0.81 (95% CI 0.66–0.89). The Siemens company reported that a low threshold of 7.7 performed accurately in excluding significant fibrosis, when the

Table 2. Calculated sensitivities and specificities at predefined cut-offs of 7.7, 9.8, 10.51, 11.3 in advanced fibrosis and their corresponding PPVs and NPVs for different prevalences using the multiple thresholds model.

Cut-off	Sensitivity	95% CI	Specificity	95% CI	Prevalence	PPV	NPV	FP*	FN*
7.70	0.93	0.82–0.98	0.34	0.13–0.65	0.05	0.07	0.99	63	0
					0.10	0.14	0.98	59	1
					0.20	0.26	0.95	53	1
					0.30	0.38	0.92	46	2
					0.40	0.49	0.88	40	3
					0.50	0.59	0.83	33	4
					0.70	0.77	0.68	20	5
9.80	0.65	0.49–0.77	0.86	0.77–0.92	0.05	0.20	0.98	13	2
					0.10	0.34	0.96	13	4
					0.20	0.54	0.91	11	7
					0.30	0.66	0.85	10	11
					0.40	0.75	0.79	8	14
					0.50	0.82	0.71	7	18
					0.70	0.91	0.51	4	25
10.51	0.51	0.31–0.70	0.93	0.85–0.96	0.05	0.26	0.97	7	2
					0.10	0.43	0.94	6	5
					0.20	0.63	0.88	6	10
					0.30	0.75	0.81	5	15
					0.40	0.82	0.74	4	20
					0.50	0.87	0.65	4	25
					0.70	0.94	0.45	2	34
11.30	0.36	0.15–0.63	0.96	0.90–0.99	0.05	0.34	0.97	4	3
					0.10	0.52	0.93	4	6
					0.20	0.71	0.86	3	13
					0.30	0.81	0.78	3	19
					0.40	0.87	0.69	2	26
					0.50	0.91	0.60	2	32
					0.70	0.96	0.39	1	45

FN, false negative; FP, false positive; NPV, negative predictive value; PPV, positive predictive value.
*Number of false positives and negatives in 100 hypothetical cases.

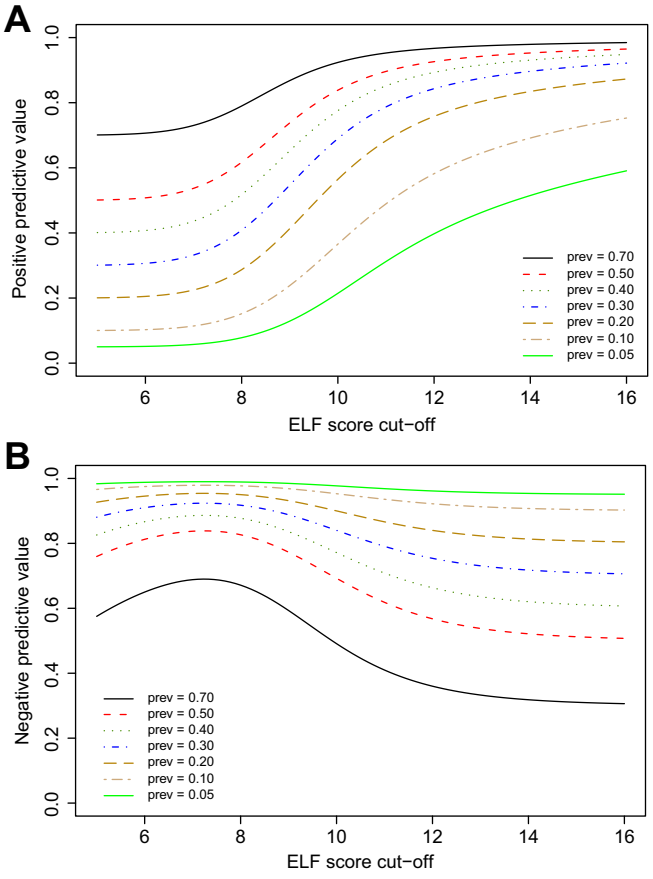


Fig. 3. Predictive values for advanced fibrosis. (A, B) These plots illustrate the corresponding (A) positive predictive values and (B) negative predictive values for different ELF cut-offs based on the multiple thresholds model using all available information for advanced fibrosis (1 color for each prevalence). ELF, enhanced liver fibrosis.

prevalence of significant fibrosis was at most 40%. In our meta-analysis, the ELF test had a sensitivity of 0.97 (95% CI 0.88–0.99) at this threshold, with high NPVs, ranging from 0.83 to 0.98 in settings with disease prevalence lower than 40% (Table S10). However, the test would not be able to exclude significant fibrosis when the prevalence was very high, as in some specialist clinical settings, where the prevalence may exceed 40%.

None of the predefined high thresholds showed high accuracy for diagnosis of significant fibrosis when the disease prevalence was lower than 30%. The high threshold recommended by the NICE guideline (10.51) would have a PPV exceeding 0.80 only in "high-prevalence" secondary or tertiary care settings, with disease prevalence more than 40% (Table S10). At lower prevalence levels such as 5% or 10%, as may be encountered in primary care and non-hepatology secondary-care settings, the PPV ranged from 0.22 to 0.66 for predefined high thresholds, while at the highest level it reached 0.66 at a threshold of 11.30.

With pre-specified high sensitivity and specificity (0.90, 0.95 and 0.98), we could evaluate other potential ELF test thresholds for diagnosing significant fibrosis, at different disease prevalence values (Table S11). At a pre-specified specificity of 0.98, the highest threshold value of 10.84 resulted in a PPV >0.80, when the prevalence was at least 30%. The plots in Fig. 5 show the corresponding PPVs and NPVs for different ELF thresholds at different pretest probabilities.

Sensitivity analysis

We conducted a sensitivity analysis to examine the impact on the meta-analytic findings of including selective patients from a population with very high disease prevalence. One study had selectively included patients with NAFLD activity score ≥3 and a majority with bridging fibrosis (F3, 31%) or cirrhosis (F4, 40%).[52] Removing this study from the meta-analysis did not significantly affect the results of the meta-analysis. The other 2 studies that were removed from our meta-analysis of advanced and significant fibrosis had long test-biopsy time intervals (>1 year). Removing these studies from the meta-analysis did not significantly affect the results of the meta-analysis either[26,34] (Tables S12–S17).

Discussion

Because of the limitations of the liver biopsy, non-invasive tests to accurately evaluate fibrosis and to assess NAFLD severity are of great interest. The ELF test has been suggested by NICE guideline as "the most cost-effective and the most appropriate test for

Table 3. Calculated predictive values based on different prevalences of advanced fibrosis with fixed sensitivities or specificities using the multiple thresholds model.

	Fixed 0.90 sensitivity				Fixed 0.95 sensitivity				Fixed 0.98 sensitivity			
Prevalence	Cut-off	Specificity	PPV	NPV	Cut-off	Specificity	PPV	NPV	Cut-off	Specificity	PPV	NPV
0.05	8.10	0.47	0.08	0.99	7.40	0.26	0.06	0.99	6.60	0.10	0.05	0.99
0.10			0.16	0.98			0.12	0.98			0.11	0.98
0.20			0.30	0.95			0.24	0.95			0.21	0.95
0.30			0.42	0.92			0.35	0.92			0.32	0.92
0.40			0.53	0.87			0.46	0.89			0.42	0.88
0.50			0.63	0.82			0.56	0.84			0.52	0.83
0.70			0.79	0.66			0.75	0.68			0.72	0.68
	Fixed 0.90 specificity				**Fixed 0.95 specificity**				**Fixed 0.98 specificity**			
Prevalence	**Cut-off**	**Sensitivity**	**PPV**	**NPV**	**Cut-off**	**Sensitivity**	**PPV**	**NPV**	**Cut-off**	**Sensitivity**	**PPV**	**NPV**
0.05	10.18	0.57	0.23	0.98	10.95	0.42	0.31	0.97	12.01	0.25	0.40	0.96
0.10			0.39	0.95			0.48	0.94			0.58	0.92
0.20			0.59	0.89			0.68	0.87			0.76	0.84
0.30			0.71	0.83			0.78	0.79			0.84	0.75
0.40			0.79	0.76			0.85	0.71			0.89	0.66
0.50			0.85	0.68			0.89	0.62			0.93	0.57
0.70			0.93	0.47			0.95	0.41			0.97	0.36

NPV, negative predictive value; PPV, positive predictive value.

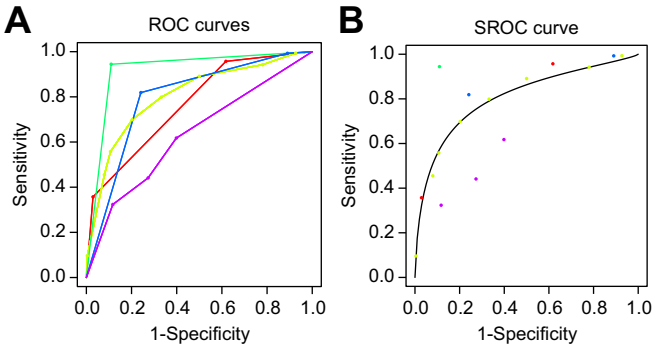


Fig. 4. Test performance for detecting significant fibrosis. (A) Multiple thresholds ROC curve and (B) Multiple thresholds SROC curve based on the multiple thresholds model using homogenized thresholds. Circles represent information on sensitivity and specificity. AUC: 0.81 (0.66, 0.89), Max Youden-index results: cut-off: 9.43, sensitivity: 0.69 (0.50, 0.83), specificity: 0.80 (0.60, 0.92). AUC, area under the ROC/SROC curve; ROC, receiver-operating characteristic; SROC, summary receiver-operating characteristic.

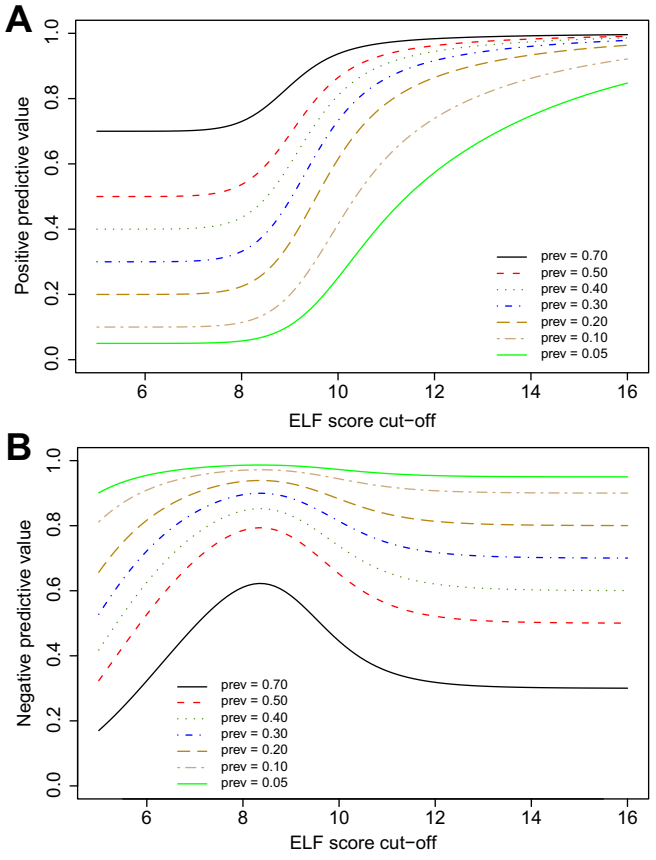


Fig. 5. Predictive values for significant fibrosis. (A, B) These plots illustrate the corresponding (A) positive predictive values and (B) negative predictive values for different ELF cut-offs based on the multiple thresholds model using all available information for significant fibrosis (1 color for each prevalence). ELF, enhanced liver fibrosis.

advanced fibrosis in adults with NAFLD".[15] However, this assertion was based on data extracted from a relatively small number of clinical studies. In the present systematic review, we synthesized the available evidence on the accuracy of this test by performing a meta-analysis of the results of 14 studies, encompassing 4,452 and 550 patients in advanced and significant fibrosis groups, respectively. Lack of sufficient data impeded the conduct of a meta-analysis for other target conditions, such as cirrhosis. The only available study of cirrhosis in 3,173 patients, reported high sensitivity of the test (83%) in excluding cirrhosis, and high specificity in detecting cirrhosis (94%), at thresholds of <9.8 and ≥11.3, respectively.

Based on the available evidence, we can conclude that the ELF test had a high sensitivity of 93% at the recommended low cut-off (7.7), however specificity is limited. At the recommended high cut-off of 9.8, specificity was 86% and even higher cut-offs would be required to achieve higher specificity. We observed high NPV, especially at low-prevalence settings such as in primary care, but, the PPV was much lower at this low disease prevalence, especially when the prevalence fell below 30%.

To minimize the risk of bias due to selective inclusion of published results, we relied on a comprehensive search strategy for finding published full texts and conference abstracts, without any restrictions. We also used 2 reviewers to independently identify the studies and extract the data, to lower the risk of errors related to single data extraction.

Because of the diversity of ELF algorithms used in the studies, we developed a new conversion formula to harmonize the results and combine multiple algorithms used to measure this biomarker. With our multiple thresholds model, we could use all thresholds reported by the included studies, without limiting ourselves to a single cut-off, as is typically done. This allowed us to evaluate the performance of the ELF test at the predefined thresholds, and to investigate new thresholds for predefined levels of sensitivity and specificity that could provide predictive values at different prevalences.

The major limitation of our systematic review and meta-analysis was the lack of information about the biopsy procedure in several individual studies. Needle gauge, the length of the biopsy and the number of pathologists who assessed the histopathological samples and their expertise were often not reported. Most included studies were conducted in tertiary centers, where the prevalence of significant and advanced fibrosis is higher than in the general population. Therefore, limited information about the performance of the test specifically derived from primary clinical settings was available for this analysis.

A number of studies have evaluated the ELF test for different purposes and in combination with other tests. However, our systematic review is focused on diagnostic performance of the ELF as a stand-alone test. We found 2 other systematic reviews with the same focus. Ooi *et al.*[56] systematically reviewed and assessed the accuracy of a few non-invasive tests that are more commonly used in clinical practice, including the ELF test. They focused on the obese population, with a BMI over 30. The prevalence of advanced fibrosis varied significantly in the included studies, from 11% to 25.7%. In their meta-analysis they included only 2 publications on the accuracy of the ELF test, with significantly different thresholds (ranged from -3.37 to 0.358 based on Guha formula, equal to 5.96 to 10.17 in Siemens scale). When comparing the results of the complex serum scores, such as the ELF test, NAFLD fibrosis score (NFS) and Fibrosis-4 score (FIB-4) with other single biomarkers in this study, they concluded that the complex panels, particularly the ELF test were more accurate.

One other related systematic review was published in 2014[57] and evaluated the performance of the ELF test to diagnose

different levels of fibrosis in patients with different chronic liver conditions, including NAFLD, viral hepatitis and primary biliary cirrhosis. Apart from the significant heterogeneity that resulted from the wide inclusion criteria, the review suffered from a failure to consider the variation of algorithms used to calculate the ELF scores, the different histological scoring systems, and the various diagnostic thresholds reported by the studies.[57]

Currently, there are no clear screening guidelines for NAFLD. However, the burden for health systems on the one hand and the risks of liver biopsy on the other, make it challenging for health care decision makers to recommend NAFLD screening in the community.[8,9] The joint EASL-EASD-EASO guideline reviewed the accuracy of a few serum markers including NFS, FIB-4, and ELF test for significant and advanced fibrosis, suggesting that "non-invasive tests may be confidently used for first-line risk stratification to exclude severe disease". The recent NICE guideline on assessment and management of NAFLD recommended ELF as an accurate biomarker and the most cost-effective test to detect advanced fibrosis.[15]

Our analysis of multiple prevalences showed that using the ELF test at a threshold of 10.51 in primary care settings with disease prevalence of 5–10% leads to very low PPV (0.26 and 0.43, respectively). However, the test can lead to PPV exceeding 0.80 in a high-prevalence setting only (>40%). Yet, at this threshold, the summary estimate of sensitivity in detecting advanced fibrosis, at 0.50, is well below the 1.00 mentioned in the NICE guideline. This may question the validity of the NICE recommendation, to "explain to people with an ELF score below 10.51 that they are unlikely to have advanced liver fibrosis".[15]

Although the available research is too limited to address biomarker accuracy in ruling out significant and advanced fibrosis in patients with NAFLD among the general population, the estimations based on our study suggest that the ELF test has a high NPV when the prevalence is lower than 30%. This highlights the value of the ELF test as a first-line test to exclude advanced fibrosis in the primary care setting and hence to avoid further evaluation by specialists. However, it is important to note that the high sensitivity of the test (>0.90) comes at the expense of limited specificity (0.30), which, given the low prevalence, means there will be a substantial number of false positive results. This needs to be considered, especially when the test is going to be applied in a clinical setting with low prevalence of the disease, as the large number of false positive results might lead patients to have unnecessary invasive and expensive procedures, like biopsy.

Due to the complexity of NAFLD and the fact that it is unlikely that a single marker would be able to accurately rule in or rule out disease, more efforts have gone into evaluating the contributions of tests.[58] In 1 study, with a before-after design, the introduction of a 2-step algorithm combining the use of FIB-4 score followed by the ELF test if required, was found to improve the detection of advanced fibrosis and cirrhosis, while reducing unnecessary referrals.[59]

ELF is now used with increasing confidence in many centers around the world, including in the US,[35] Japan,[25] South Korea,[60] the UK[24,34] and many other European countries.[26–31,33] However, the availability of the test in clinics varies from country to country, due to different regulatory requirements and insurance coverage. Moreover, the test is being used in clinical settings that differ in the prevalence of advanced fibrosis, ranging from 5–10% in the general population to more than 30% or 40% in secondary and tertiary referral centers.[61] Therefore, the prevalence of advanced fibrosis in the target population should be carefully considered when selecting the desired test positivity threshold. Fig. 3 and 5 in the current manuscript will assist clinicians to identify suitable thresholds that are tailored to their specific clinical setting/prevalence levels and the balance of sensitivity/specificity they require.

Conclusions

The meta-analysis of the available evidence showed ELF exhibits high sensitivity but limited specificity to exclude advanced and significant fibrosis in patients with NAFLD, when using the low-cut-off scores. The performance of the test for diagnosing significant and advanced fibrosis at higher score thresholds was also found to be limited in the context of low disease prevalence (5–10%). These estimations were, however, projected based on the observed cumulative distribution of the ELF in diseased and non-diseased populations in the primary studies conducted in higher prevalence settings. Clinicians should carefully consider the likely *a priori* disease prevalence in their clinical practice setting and select a suitable test threshold locally, to achieve the desired level of test performance, in terms of sensitivity, specificity and predictive values. Further comparative studies of high methodological quality are necessary to obtain more reliable evidence of accuracy of the ELF test, especially in different clinical settings.

Abbreviations

ARFI, acoustic radiation force impulse imaging; AUC, area under the ROC/SROC curve; CRN, Clinical Research Network; ELF, enhanced liver fibrosis; FIB-4, fibrosis-4; HA, hyaluronic acid; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; NFS, NAFLD fibrosis score; NPV, negative predictive value; PIIINP, type III procollagen peptide; PPV, positive predictive value; ROC, receiver-operating characteristic; SROC, summary receiver-operating characteristic; TIMP, tissue inhibitor of metalloproteinase-1.

Financial support

This systematic review has been conducted as part of the evidence synthesis efforts in the LITMUS (Liver Investigation: Testing Marker Utility in Steatohepatitis) study. The LITMUS study is funded by the Innovative Medicines Initiative (IMI2) Program of the European Union (Grant Agreement 777377). The funder and the authors' institutions had no role in the development of this systematic review.

Conflict of interest

QMA is coordinator of the IMI2 LITMUS consortium. He reports research grant funding from Abbvie, Allergan/Tobira, AstraZeneca, GlaxoSmithKline, Glympse Bio, Novartis Pharma AG, Pfizer Ltd., Vertex; consultancy on behalf of Newcastle University for Abbott Laboratories, Acuitas Medical, Allergan/Tobira, Blade, BNN Cardio, Cirius, CymaBay, EcoR1, E3Bio, Eli Lilly & Company Ltd., Galmed, Genfit SA, Gilead, Grunthal, HistoIndex, Indalo, Imperial Innovations, Intercept Pharma Europe Ltd., Inventiva, IQVIA, Janssen, Kenes, Madrigal, MedImmune, Metacrine, NewGene, NGMBio, North Sea Therapeutics, Novartis, Novo Nordisk A/S, Pfizer Ltd., Poxel, ProSciento, Raptor Pharma, Servier, Viking Therapeutics; and speaker fees from Abbott Laboratories,

Allergan/Tobira, BMS, Clinical Care Options, Falk, Fishawack, Genfit SA, Gilead, Integritas Communications, MedScape.

MJB is a share-holder and employee of Pfizer.

Please refer to the accompanying ICMJE disclosure forms for further details.

Authors' contributions

YV, MHZ, PB, MJB, QMA and JL contributed in designing the study and YV prepared the draft of the protocol. Search strategy has been developed by YV, RS, and MHZ with great contribution of ZB. YV and JL screened the references resulted from the literature search and extracted the required data. Statistical analyses and interpretation have been performed mainly by YV and MHZ. JB and JLö conducted the analysis of harmonization of algorithms in different data sets and other authors contributed in interpretation of the results. All authors reviewed and critically revised the protocol and the manuscript.

†The LITMUS systematic review team

Michael Pavlides (Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, UK; Translational Gastroenterology Unit, University of Oxford, UK; Oxford NIHR Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust and the University of Oxford, Oxford, UK); Christina Levick (Translational Gastroenterology Unit, University of Oxford, UK); Kevin Duffin (Eli Lilly and Company Ltd (LLY), USA); Craig Hyde (Pfizer Worldwide Research and Development (PFE), 1 Portland Square, Cambridge, USA); Tim Bauer (SomaLogic, Inc. USA); Pierre Bedossa (University Paris-Diderot, Pathology, Hôpital Beaujon, APHP (APHP), Paris, France); Diana Leeming (Nordic Bioscience (NB), Denmark); Ann Daly (Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, UK); Remy Hanf (Genfit SA, France); Pablo Ortiz (OWL/metabolomics, One Way Liver S.L, Spain); Matej Oresic (Örebro University (ORU), Sweden); Detlef Schuppan (University Medical Center Mainz (UMCM), Germany); Guido Hanauer (Takeda Development Center Europe, London, UK); Yu Chen (Eli Lilly and Company Ltd (LLY), USA); Elizabeth Shumbayawonda (Perspectum Diagnostics Ltd., Oxford, UK), Peter Nissen Bjerring (Novo Nordisk, Bagsværd, Denmark), Koos Zwinderman (Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam UMC, University of Amsterdam, The Netherlands).

Registration

The protocol of the LITMUS systematic review is available in PROSPERO: CRD42018106821 (https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=106821).

Acknowledgment

The authors sincerely thank the following investigators for kindly sharing additional data from their studies, which enabled this meta-analysis to be performed:

Professor Christopher D. Byrne and Dr. Joshua A. Welsh, (Human Development and Health Academic Unit, Faculty of Medicine, University of Southampton, Southampton, United Kingdom), Dr. Rudolf Stauber (Institute of Pathology, Graz, Austria), Dr. Stephen Djedjos (Gilead Sciences, Inc.), Dr. Peter Eddowes (Centre for Liver Research, University of Birmingham, Birmingham, United Kingdom).

This work has been funded by the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No. 777377 (LITMUS). This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhep.2020.03.036.

References

[1] Adams LA, Anstee QM, Tilg H, Targher G. Non-alcoholic fatty liver disease and its relationship with cardiovascular disease and other extrahepatic diseases. Gut 2017;66(6):1138–1153.
[2] Hadi A, Mohammadi H, Miraghajani M, Ghaedi E. Efficacy of synbiotic supplementation in patients with nonalcoholic fatty liver disease: a systematic review and meta-analysis of clinical trials: synbiotic supplementation and NAFLD. Crit Rev Food Sci Nutr 2019;59:2494–2505.
[3] Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology 2016;64(1):73–84.
[4] Younossi ZM, Stepanova M, Afendy M, Fang Y, Younossi Y, Mir H, et al. Changes in the prevalence of the most common causes of chronic liver diseases in the United States from 1988 to 2008. Clin Gastroenterol Hepatol 2011;9(6):524–530.e1. quiz e60.
[5] Mann JP, De Vito R, Mosca A, Alisi A, Armstrong MJ, Raponi M, et al. Portal inflammation is independently associated with fibrosis and metabolic syndrome in pediatric nonalcoholic fatty liver disease. Hepatology 2016;63(3):745–753.
[6] Dulai PS, Singh S, Patel J, Soni M, Prokop LJ, Younossi Z, et al. Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: systematic review and meta-analysis. Hepatology 2017;65(5):1557–1565.
[7] Singh S, Allen AM, Wang Z, Prokop LJ, Murad MH, Loomba R. Fibrosis progression in nonalcoholic fatty liver vs nonalcoholic steatohepatitis: a systematic review and meta-analysis of paired-biopsy studies. Clin Gastroenterol Hepatol 2015;13(4):643–654.e9.
[8] Liver EAftSot. EASL-ALEH clinical practice guidelines: non-invasive tests for evaluation of liver disease severity and prognosis. J Hepatol 2015;63(1):237–264.
[9] Liver EAftSot, Diabetes EAftSo. EASL-EASD-EASO clinical practice guidelines for the management of non-alcoholic fatty liver disease. Obes Facts 2016;9(2):65–90.
[10] Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, et al. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American Association for the Study of Liver Diseases. Hepatology 2018;67(1):328–357.
[11] Rinella ME, Tacke F, Sanyal AJ, Anstee QM, participants of the AEW. Report on the AASLD/EASL joint workshop on clinical trial endpoints in NAFLD. J Hepatol 2019;70:1424–1436.
[12] The Royal College of Pathologists. Tissue Pathways for Liver Biopsies for the Investigation of Medical Disease and for Focal Lesions. London: The Royal College of Pathologists; 2008.
[13] Rockey DC, Caldwell SH, Goodman ZD, Nelson RC, Smith AD. Liver biopsy. Hepatology 2009;49(3):1017–1044.
[14] Ratziu V, Charlotte F, Heurtier A, Gombert S, Giral P, Bruckert E, et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. Gastroenterology 2005;128(7):1898–1906.
[15] NICE guideline NCGC. Non-alcoholic Fatty Liver Disease. Assessment and Management of Non-alcoholic Fatty Liver Disease (NAFLD) in Adults, Children and Young People. London: Guideline National Institute for Health and Care Excellence; 2015.
[16] Rosenberg WM, Voelker M, Thiel R, Becka M, Burt A, Schuppan D, et al. Serum markers detect the presence of liver fibrosis: a cohort study. Gastroenterology 2004;127(6):1704–1713.
[17] Guha IN, Parkes J, Roderick P, Chattopadhyay D, Cross R, Harris S, et al. Noninvasive markers of fibrosis in nonalcoholic fatty liver disease: validating the European liver fibrosis panel and exploring simple markers. Hepatology 2008;47(2):455–460.
[18] Healthcare S. ELF Physician Brochure. Germany: Siemens Healthcare, Laboratory Diagnostics; 2016.
[19] Glen J, Floros L, Day C, Pryke R. Non-alcoholic fatty liver disease (NAFLD): summary of NICE guidance. BMJ 2016;354:i4428.

[20] Nobili V, Parkes J, Bottazzo G, Marcellini M, Cross R, Newman D, et al. Performance of ELF serum markers in predicting fibrosis stage in pediatric non-alcoholic fatty liver disease. Gastroenterology 2009;136(1):160–167.
[21] Anstee QM, Lawitz EJ, Alkhouri N, Wong VWS, Romero-Gomez M, Okanoue T, et al. Routinely available noninvasive tests discriminate advanced fibrosis due to NASH in the phase 3 stellar trials of the ASK1 inhibitor selonsertib. Hepatology 2018;68(Supplement 1):953A.
[22] Boursier J, Guillaume M, Moal V, Delabaudiere C, Zuberbuhler F, Robic MA, et al. First direct comparison in NAFLD of the blood fibrosis tests fibrometer and ELF. Hepatology 2018;68(Supplement 1):88A–89A.
[23] Dvorak K, Stritesky J, Petrtyl J, Vitek L, Sroubkova R, Lenicek M, 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(10):e111551.
[24] Eddowes PJ, McDonald N, Davies N, Semple SIK, Kendall TJ, Hodson J, et al. Utility and cost evaluation of multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease. Aliment Pharmacol Ther 2017;47(5):631–644.
[25] Itoh Y, Seko Y, Shima T, Nakajima T, Mizuno K, Kawamura Y, et al. Accuracy of non-invasive scoring systems for diagnosing non-alcoholic steatohepatitis-related fibrosis: Multicenter validation study. Hepatol Res 2018;48(13):1099–1107.
[26] Karlas T, Dietrich A, Peter V, Wittekind C, Lichtinghagen R, Garnov N, et al. Evaluation of Transient Elastography, acoustic radiation force impulse imaging (ARFI), and enhanced liver function (ELF) score for detection of fibrosis in morbidly obese patients. PLoS One 2015;10(11):e0141649.
[27] Lopez IC, Aroca FG, Bernal MDF, Mompean JAL, Bernal AB, Martinez AMH, 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(9):2347–2353.
[28] Lykiardopoulos B, Hagstrom H, Fredrikson M, Ignatova S, Stal P, Hultcrantz R, 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(12):e0167776.
[29] Miele L, De Michele T, Marrone G, Antonietta Isgro M, Basile U, Cefalo C, 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(4):e397–e402.
[30] Polyzos SA, Slavakis A, Koumerkeridis G, Katsinelos P, Kountouras J. Noninvasive liver fibrosis tests in patients with nonalcoholic fatty liver disease: an external validation cohort. Horm Metab Res 2019;51(2):134–140.
[31] Sanyal AJ, Harrison SA, Cordonnier G, Brozek J, Roudot A, Praca E, et al. Assessment of serum levels of Chitinase-3-like protein 1 (CHI3L1) improves identification of the NASH patients at risk who should be treated. Hepatology conference: 67th Annual Meeting of the American association for the study of liver diseases: the liver Meeting 2016 United States conference Start: 20161111 conference End: 20161115 2016; 2016;63(1):327A.
[32] Schulze Pellengahr S, Neuber S, Wagner-Golbs A, Christiansen N, Schatz P, Kamlage B, et al. The combination of an androgen, a bile acid and an eicosanoid is a promising biomarker for liver fibrosis in non-alcoholic fatty liver disease patients. United Eur Gastroenterol J 2018;6(8 Supplement):A545.
[33] Staufer K, Marculescu R, Obermayer-Pietsch B, Stift J, Lackner C, Trauner M, et al. Enhanced liver fibrosis (ELF) score accurately detects advanced fibrosis in nonalcoholic fatty liver disease (NAFLD). Z Gastroenterologie 2018;56(5):e41.
[34] Welsh JA, Scorletti E, Clough GF, Englyst NA, Byrne CD. Leukocyte extracellular vesicle concentration is inversely associated with liver fibrosis severity in NAFLD. J Leukoc Biol 2018;104(3):631–639.
[35] Younossi ZM, Page S, Rafiq N, Birerdinc A, Stepanova M, Hossain N, et al. A biomarker panel for non-alcoholic steatohepatitis (NASH) and NASH-related fibrosis. Obes Surg 2011;21(4):431–439.
[36] Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 2005;41(6):1313–1321.
[37] Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. The French METAVIR Cooperative Study Group. Hepatology 1994;20(1 Pt 1):15–20.
[38] Batts KP, Ludwig J. Chronic hepatitis. An update on terminology and reporting. Am J Surg Pathol 1995;19(12):1409–1417.
[39] Bedossa P, Arola J, Susan D, Gouw A, Maria G, Lackner K, et al. The EPoS staging system is a reproducible 7-tierfibrosis score for NAFLD adapted both to glass slides and digitized images (e-slides). J Hepatol 2018;68:S553.
[40] Desmet VJ, Gerber M, Hoofnagle JH, Manns M, Scheuer PJ. Classification of chronic hepatitis: diagnosis, grading and staging. Hepatology 1994;19(6):1513–1520.
[41] Ishak K, Baptista A, Bianchi L, Callea F, De Groote J, Gudat F, et al. Histological grading and staging of chronic hepatitis. J Hepatol 1995;22(6):696–699.
[42] Knodell RG, Ishak KG, Black WC, Chen TS, Craig R, Kaplowitz N, et al. Formulation and application of a numerical scoring system for assessing histological activity in asymptomatic chronic active hepatitis. Hepatology 1981;1(5):431–435.
[43] Scheuer PJ. Classification of chronic viral hepatitis: a need for reassessment. J Hepatol 1991;13(3):372–374.
[44] Bedossa P, Poitou C, Veyrie N, Bouillot JL, Basdevant A, Paradis V, et al. Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. Hepatology 2012;56(5):1751–1759.
[45] Brunt EM, Janney CG, Di Bisceglie AM, Neuschwander-Tetri BA, Bacon BR. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. Am J Gastroenterol 1999;94(9):2467.
[46] Matteoni CA, Younossi ZM, Gramlich T, Boparai N, Liu YC, McCullough AJ. Nonalcoholic fatty liver disease: a spectrum of clinical and pathological severity. Gastroenterology 1999;116(6):1413–1419.
[47] Younossi ZM, Stepanova M, Rafiq N, Makhlouf H, Younoszai Z, Agrawal R, et al. Pathologic criteria for nonalcoholic steatohepatitis: interprotocol agreement and ability to predict liver-related mortality. Hepatology 2011;53(6):1874–1882.
[48] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155(8):529–536.
[49] Schneider A, Linde K, Reitsma JB, Steinhauser S, Rücker G. A novel statistical model for analyzing data of a systematic review generates optimal cut-off values for fractional exhaled nitric oxide for asthma diagnosis. J Clin Epidemiol 2017;92:69–78.
[50] Steinhauser S, Schumacher M, Rücker G. Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. BMC Med Res Methodol 2016;16(1):97.
[51] Boursier J, Guillaume M, Leroy V, Irlès M, Roux M, Lannes A, et al. New sequential combinations of non-invasive fibrosis tests provide an accurate diagnosis of advanced fibrosis in NAFLD. J Hepatol 2019;71(2):389–396.
[52] Anstee QM, Lawitz EJ, Alkhouri N, Wong VWS, Romero-Gomez M, Okanoue T, et al. Noninvasive tests accurately identify advanced fibrosis due to NASH: baseline data from the STELLAR trials. Hepatology 2019;70:1521–1530.
[53] Bürkner PC, Doebler P. Testing for publication bias in diagnostic meta-analysis: a simulation study. Stat Med 2014;33(18):3061–3077.
[54] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151(4):264–269.
[55] Staufer K, Halilbasic E, Spindelboeck W, Eilenberg M, Prager G, Stadlbauer V, 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(8):1113–1123.
[56] Ooi GJ, Mgaieth S, Eslick GD, Burton PR, Kemp WW, Roberts SK, et al. Systematic review and meta-analysis: non-invasive detection of non-alcoholic fatty liver disease related fibrosis in the obese. Obes Rev 2018;19(2):281–294.
[57] 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(4):e92772.
[58] Verhaegh P, Bavalia R, Winkens B, Masclee A, Jonkers D, Koek G. Noninvasive tests Do not accurately Differentiate nonalcoholic steatohepatitis from simple steatosis: a systematic review and meta-analysis. Clin Gastroenterol Hepatol 2018;16(6):837–861.
[59] Srivastava A, Gailer R, Tanwar S, Trembling P, Parkes J, Rodger A, et al. Prospective evaluation of a primary care referral pathway for patients with non-alcoholic fatty liver disease. J Hepatol 2019;71:371–378.
[60] Kim BK, Kim HS, Park JY, Ahn SH, Chon CY, Park YN, 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(7):e41964.
[61] Koehler EM, Plompen EP, Schouten JN, Hansen BE, Darwish Murad S, Taimr P, et al. Presence of diabetes mellitus and steatosis is associated with liver stiffness in a general population: the Rotterdam study. Hepatology 2016;63(1):138–147.