ORIGINAL ARTICLE

Liver INTERNATIONAL | WILEY

Prognostic accuracy of FIB-4, NAFLD fibrosis score and APRI for NAFLD-related events: A systematic review

Jenny Lee[1] 🆔 | Yasaman Vali[1] 🆔 | Jerome Boursier[2,3] | Rene Spijker[4,5] | Quentin M. Anstee[6,7] 🆔 | Patrick M. Bossuyt[1] | Mohammad H. Zafarmand[1]

[1]Epidemiology and Data Science, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

[2]Hepato-Gastroenterology Department, Angers University Hospital, Angers, France

[3]HIFIH Laboratory, UPRES EA3859, Angers University, Angers, France

[4]Medical Library, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

[5]Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

[6]The Newcastle Liver Research Group, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

[7]Newcastle NIHR Biomedical Research Centre, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

Correspondence
Jenny Lee, Epidemiology and Data Science, Amsterdam UMC, Location AMC, Meibergdreef 9, 1105AZ Amsterdam, the Netherlands.
Email: j.a.lee@amsterdamumc.nl

Abstract

Background & Aims: Fibrosis is the strongest predictor for long-term clinical outcomes among patients with non-alcoholic fatty liver disease (NAFLD). There is growing interest in employing non-invasive methods for risk stratification based on prognosis. FIB-4, NFS and APRI are models commonly used for detecting fibrosis among NAFLD patients. We aimed to synthesize existing literature on the ability of these models in prognosticating NAFLD-related events.

Methods: A sensitive search was conducted in two medical databases to retrieve studies evaluating the prognostic accuracy of FIB-4, NFS and APRI among NAFLD patients. Target events were change in fibrosis, liver-related event and mortality. Two reviewers independently performed reference screening, data extraction and quality assessment (QUAPAS tool).

Results: A total of 13 studies (FIB-4:12, NFS: 11, APRI: 10), published between 2013 and 2019, were retrieved. All studies were conducted in a secondary or tertiary care setting, with follow-up ranging from 1 to 20 years. All three markers showed consistently good prognostication of liver-related events (AUC from 0.69 to 0.92). For mortality, FIB-4 (AUC of 0.67-0.82) and NFS (AUC of 0.70-0.83) outperformed APRI (AUC of 0.52-0.73) in all studies. All markers had inconsistent performance for predicting change in fibrosis stage.

Conclusions: FIB-4, NFS, and APRI have demonstrated ability to risk stratify patients for liver-related morbidity and mortality, with comparable performance to a liver biopsy, although more head-to-head studies are needed to validate this. More refined models to prognosticate NAFLD-events may further enhance performance and clinical utility of non-invasive markers.

- - - - - - - - - -

1 | INTRODUCTION

In the next 20 years, non-alcoholic fatty liver disease (NAFLD) is projected to become the leading cause of liver transplantation.[1,2] The global prevalence of NAFLD is approximately 25%, among which a proportion may progress to develop non-alcoholic steatohepatitis (NASH).[3] The prevalence of NAFLD-related cirrhosis as the underlying disease among patients undergoing liver transplantation for hepatocellular carcinoma (HCC) has markedly increased in Europe and the United States.[4,5] Patients with NASH have a higher risk of progression to liver fibrosis,[6,7] and those with advanced fibrosis or cirrhosis trend towards more complications of liver failure and HCC compared to those without fibrosis.[8]

Liver fibrosis is considered the strongest predictor for long-term clinical outcomes in NAFLD patients.[9] Accurate assessment of NASH or fibrosis stage is resource intensive and error-prone, as a liver biopsy is currently required to confirm the diagnosis.[10,11] Moreover, biopsies carry risks for the patient such as severe complications and pain, leaving many unwilling to undergo this invasive procedure.

There is growing promise in risk stratification using non-invasive markers of NAFLD for identifying patients more likely to develop severe liver events. Using markers that are more reliable than a biopsy would circumvent the limitations of a biopsy in stratifying patients. Optimally performing prognostic markers can eventually replace a biopsy and aid clinical decision-making, as well as facilitate recruitment of patients more likely to benefit from participation in clinical trials.

Simple non-invasive panels such as the NAFLD Fibrosis Score (NFS) and Fibrosis-4 (FIB-4) are recommended by the EASL-EASD-EASO Clinical Practice Guidelines as part of the diagnostic regimen for ruling out advanced fibrosis.[12] The guidelines further recommend the use of NFS and FIB-4 as prognostic markers to rule out progression to severe disease, including liver-related and all-cause mortality. Other multimarker models such as the aspartate aminotransferase (AST)/platelet ratio index (APRI) are also used for fibrosis staging and prediction of liver-related events.[13] Reviewing the literature, we found other markers such as Enhanced Liver Fibrosis (ELF) test or FibroScan had limited assessment for their prognostic ability.

Despite established diagnostic performance, there is limited understanding of the relative merits of the prognostic ability of non-invasive NAFLD markers, and their comparability to a liver biopsy. While many studies have assessed diagnostic performance of these markers in reference to a biopsy, more convincing evidence would link these markers to future clinical events. In this context, we aimed to conduct a systematic review of studies on the accuracy of FIB-4,

Key Points

- FIB-4, NFS and APRI showed consistently good ability to prognosis future occurrence of liver-related events among adults with NAFLD.
- FIB-4 and NFS outperformed APRI in prognosticating mortality.
- In clinical practice, FIB-4 and NFS can be used serially to monitor disease progression and improve risk stratification.
- Direct comparisons showed promising ability of non-invasive markers to risk stratify patients with some studies concluding comparable performance to a liver biopsy.
- All three markers had inconsistent accuracy for predicting change in fibrosis stage.

NFS and APRI in prognosis of fibrosis progression, and liver-related events including mortality.

2 | METHODS

This systematic review was conducted as part of the evidence synthesis efforts of the LITMUS project (Liver Investigation: Testing Marker Utility in Steatohepatitis), funded by the European Union's IMI2 program. LITMUS aims to evaluate biomarkers for drug development in NAFLD. The protocol of the complete systematic review is available in PROSPERO (registration number: CRD42019136118). This study report was prepared using the PRISMA-DTA statement (Table S1).

2.1 | Search strategy

A sensitive search strategy, containing words in the title/abstract or text words across the record and the medical subject heading (MeSH), was developed in close collaboration with an experienced information specialist (RS). The full search strategy is available in Table S2. MEDLINE (via OVID) and EMBASE (via OVID) were searched to retrieve potentially eligible studies from inception to June 2019. A search update was conducted in June 2020. Additionally, we manually screened reference lists and contacted partners within the LITMUS consortium.

2.2 | Study selection

Search results of the two databases were merged and deduplicated using Endnote. Title and abstracts were screened by two independent reviewers (JL and YV), using Rayyan QCRI (http://rayyan.qcri.org). Full texts of potentially eligible studies were retrieved for evaluation against a pre-specified inclusion criterion by the same two reviewers. Any discrepancies were resolved by discussion.

2.3 | Inclusion and exclusion criteria

We searched for studies published in peer-reviewed journals that had assessed the prognostic accuracy of at least one of the biomarkers of interest (FIB-4, NFS, APRI) in predicting future liver-related events, or changes in fibrosis stage at future biopsies. Publications in any language were eligible for inclusion.

Studies that included adults (≥18 years) diagnosed (based on liver histology) or clinically suspected with NAFLD, and data on either FIB-4, NFS, or APRI were eligible. Studies in a mixed cohort of conditions (eg NAFLD and viral hepatitis patients) were only included if outcomes were separately reported for NAFLD patients.

The target events of interest were the following:

- worsening (or improvement) of fibrosis stage, evaluated preferably by using the NASH CRN score[14] and the EPoS staging system[15] for all stages of fibrosis or any dichotomized fibrosis status (eg F0-F2 vs F3-F4);
- other liver-related outcomes of interest, including model of end stage liver disease (MELD) score ≥15; liver transplant; HCC; large oesophageal/gastric varices; ascites; increase in hepatic venous pressure gradient (HVPG) >10 mm Hg; histological progression to cirrhosis; hospitalization (as defined by a stay of ≥24 hours) for onset of: variceal bleed, hepatic encephalopathy, spontaneous bacterial peritonitis;
- mortality (liver-related or all-cause).

Studies that reported the area under the ROC curve (AUC) or Harrell's C index for expressing the prognostic performance in predicting changes in fibrosis stage, liver-related events of interest or mortality were included. Studies reporting only measures of association, such as a relative risk, hazard ratio, odds ratio or standard deviation of change, without a direct measure of classification, were excluded.

2.4 | Data extraction and quality assessment

The following data were extracted from each included study: study characteristics, clinical characteristics, index test features, target event features (if applicable) and overall performance of the test in terms of AUC or C index. Data were independently extracted and cross-checked by a second reviewer (JL and YV).

The Quality Assessment of Prognostic Accuracy Studies (QUAPAS) tool was used to assess the methodological quality and risk of bias in the included studies.[16] In short, QUAPAS is a modification of the existing Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool,[17] revised to account for items unique to prognostic accuracy study designs. QUAPAS follows the same domain-based framework as QUADAS-2. Two independent reviewers (JL and YV) evaluated risk of bias and concerns for applicability using the five domains (participant recruitment, index test, target event, study flow, analysis), assigning each study with a judgement of 'low', 'high', or 'unclear' risk. See Table S3 for the QUAPAS tool.

2.5 | Statistical Analysis

Given the anticipated heterogeneity between studies, a meta-analysis was not considered.

3 | RESULTS

3.1 | Search results

Following deduplication, 4510 studies were eligible for title and abstract screening, of which 126 full texts were screened. We excluded 114 studies in this phase, following the inclusion and exclusion criteria. Two studies that were identified during the search update, despite having prognostic accuracy data, did not present enough data for inclusion.[18,19] Finally, a total of 13 studies, published between 2013 and 2019, were included in the present systematic review (Figure 1).

3.2 | Characteristics of included studies

The majority of studies (12/13) were comparative accuracy studies, in which two or more biomarkers were evaluated within the same cohort for a given target event. Twelve studies were identified for FIB-4, 11 for NFS and 10 for APRI. The study group consisted of NASH patients in three studies,[20-22] NAFLD-cirrhotic patients in one study,[23] and all others were NAFLD patients. All studies were conducted in a secondary or tertiary care setting. At baseline, the prevalence of diabetic patients ranged from 9% to 78% and hypertension from 11% to 55%. Mean body mass index (BMI) spanned from 28 to 35 kg/m². Characteristics of the included studies are summarized in Table 1.

3.3 | Quality assessment

The overall risk of bias and applicability concerns are summarized in Figure 2. In short, one study had unclear risk of bias in the participant recruitment domain because of sparsely reported enrolment
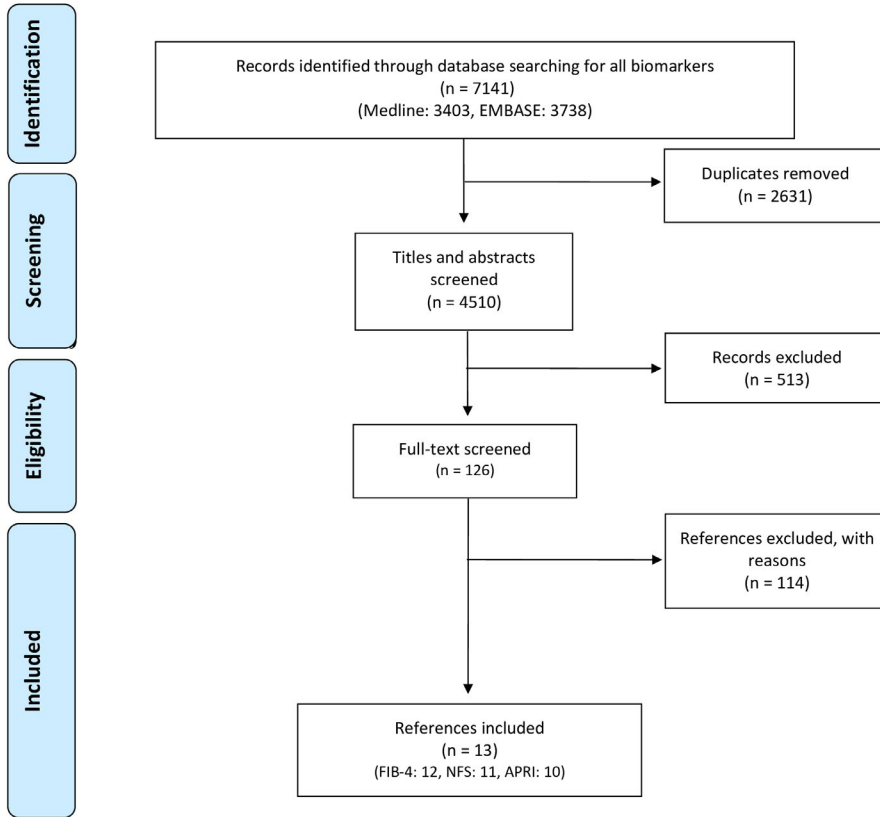
FIGURE 1 Flow diagram of included studies

or exclusion criteria.[13] Four studies had high applicability concerns for including only NASH patients[20-22] or cirrhotic-NAFLD patients.[23] Under the index test domain, four studies were graded as unclear risk of bias as use of a pre-specified threshold was not reported[20,22-24] and four studies had high applicability concerns for variability in the APRI formula (upper limit of normal for AST heterogeneous).[13,24-26] Only one study had high risk of bias in the target event domain as the outcome for the study was determined by interviews.[24] Eleven studies had unclear risk of bias in the study flow domain as information on the target event was not available for all participants, and the relationship between loss to follow-up and the index tests was not explored. Lastly, four studies were graded at high risk of bias for failing to apply methods to account for censoring and competing events.[13,20,22,27] Only one study had low risk of bias in the analysis domain.[23]

3.4 | Prognosis of change in fibrosis stage

Table 2 shows the AUC or C-index for the studies included in this systematic review. Change in fibrosis stage (fibrosis progression or regression) was evaluated as the event of interest in three studies.[13,20,22] All three studies assessed the ability of FIB-4, NFS and APRI for prognosis of fibrosis progression, defined as an increase of at least one point in fibrosis score. Two studies looked at progression into advanced fibrosis (F ≥ 3),[13,28] and another at fibrosis regression (decrease of at least one point in fibrosis score).[22] The cumulative incidence (number of study participants with the target event relative

to all study participants at the start of the observation period) of fibrosis spanned from 16% to 43%, with a mean follow-up period of 1-6.6 years.

For FIB-4, the prognostic accuracy for fibrosis progression including progression to advanced fibrosis ranged from an AUC of 0.65 (0.54-0.76) to 0.81 (0.73-0.89). The AUC for NFS ranged from 0.65 (0.56-0.73) to 0.83 (0.74-0.92), and for APRI from 0.65 (0.53-0.73) to 0.72 (0.65-0.80).

Few studies reported details regarding threshold values and corresponding sensitivity and specificity. One study used a threshold of 0.2 for all three markers.[20] For NFS, suggested high and low thresholds of 0.676 (Se: 0.28, Sp: 0.9) and −1.455 (Se: 0.91, Sp: 0.46), respectively, were used in one study.[29] One study also reported sensitivity and specificity data, but with no reporting of threshold.[13]

3.5 | Prognosis of liver-related events

Six studies evaluated liver-related events among NAFLD patients.[21,23,25,30-32] Liver-related events were defined as a combination of clinical outcomes, consisting of but not limited to ascites, esophageal varices, encephalopathy, variceal bleeding, decompensated liver disease, HCC and liver transplantation. Each study assessed a different cluster of events (see Table 2 for details). Two studies included more severe clinical outcomes such as liver failure or death.[21,30] One study evaluated solely HCC.[23] The mean follow-up was 1.9-19.9 years, with cumulative incidence ranging from 6% to 56%.

TABLE 1 Characteristics of the included studies

Author	No. of centres	Country(s)	N	Females, n (%)	Mean Age, ±SD	Mean BMI (kg/m²), ±SD	Mean ALT (IU/L), ±SD	Mean AST (IU/L), ±SD	Comorbidities, n (%)	
									DM	HTN
Angulo (2013)[25]	7	USA, Australia, UK, Iceland, Thailand, Italy	309	182 (57)	52 (43-61)	33 (29.4-36)	61 (38-85)	50 (37-78)	116 (36)	152 (48)
Treeprasertsuk (2013)[27]	NR	USA	302	169 (56)	47 ± 13	33.6 ± 6.2	61.5 ± 43.3	41.4 ± 21.9	48 (16)	124 (41)
Xun (2014)[24]	1	China	180	84 (47)	39 (30-49)	26.0 ± 3.1	129 ± 103	83.7 ± 98.2	17 (9)	20 (11)
Sebastiani (2015)[21a]	1	Canada	148	55 (30)	49.5 ± 10.5	31.3 ± 5.4	NR	NR	49 (33)	58 (39)
McPherson (2015)[28]	1	UK	108	48 (44)	48 ± 12	33.9 ± 5.0	112 ± 80	73 ± 48	52 (48)	NR
Boursier (2016)[26]	1	France	360	124 (34)	59.3 ± 14.3	NR	50 ± 45	40 ± 33	NR	NR
Vilar-Gomez (2017)[22a]	1	Cuba	261	159 (61)	48.5 ± 9.6	31.3 ± 5.3	52.4 ± 34.5	35.2 ± 20.7	90 (35)	NR
Chalasani (2018)[20a]	8	NR	191	NR	NR	NR	NR	NR	NR	NR
Peleg (2018)[32]	1	Israel	153	85 (56)	49.5	NR	NR	NR	97 (63)	64 (41)
Ioannou (2019)[23b]	1221	USA	7068	318 (4.5)	67.1 ± 9.7	33.0 ± 6.6	NR	NR	5506 (78)	NR
Siddiqui (2019)[13]	NR	NR	292	186 (64)	48.9 ± 11.7	34.7 ± 6.3	75.8 ± 50.5	53.9 ± 36.8	113 (39)	160 (55)
Onnerhag (2019)[31]	1	Sweden	144	61 (42.4)	53.2 ± 13.4	28.0 ± 4.6	79.1 ± 64.5	51.9 ± 41.8	32 (22)	66 (46)
Hagstrom (2019)[30]	2	Sweden	646	244 (38)	50 (38-58)	28.0 (25.7-30.8)	73 (49-106)	40 (31-59)	93 (14)	196 (30)

Abbreviations: DM, diabetes mellitus; HTN: hypertension; NR, not reported.

[a]Non-alcoholic steatohepatitis patients.

[b]NAFLD-cirrhotic patients.

FIGURE 2 Graphical summary of the risk of bias and applicability concerns of the included studies using the QUAPAS tool

The AUC for prognosis of liver-related events ranged from 0.71 to 0.89 for FIB-4, 0.72-0.92 for NFS, and 0.69-0.89 (0.82-0.96) for APRI (Table 2). In the two studies that conducted statistical testing, both showed significant differences ($P < .005$) between the three markers and the null hypothesis (AUC of 0.5).[25,31] In one study that compared non-invasive methods to a liver biopsy, FIB-4 and APRI had higher AUC than histological fibrosis.[21] Length of follow-up period did not seem to influence the performance of any biomarker in a consistent pattern.

In prognosticating liver-related events, most studies reported using either one or both the suggested high and low thresholds for NFS (low: −1.45, high: 0.676) and APRI (low: 0.5, high: 1.5). For FIB-4, two studies used a single threshold of 3.25 (one study finding a sensitivity and specificity of 0.59 and 0.92, respectively),[21,23] while the rest adhered to the suggested low threshold of 1.3 and/or high threshold of 2.67. In the sole study that reported paired point accuracy data, the high threshold showed a sensitivity and specificity of 0.50 and 0.90 for NFS, and 0.50 and 0.92, for APRI respectively.[21]

3.6 | Prognosis of mortality (liver-related and all-cause)

All-cause mortality was the most frequently investigated event, evaluated in seven studies.[24-27,30-32] One study additionally looked at liver-related mortality.[26] The cumulative incidence was between 5% and 59%; mean follow-up and ranged from 1.9 to 19.9 years.

The prognostic accuracy of FIB-4, expressed as the AUC, ranged from 0.67 (0.58-0.76) to 0.82 (0.75-0.90) (Table 2). The AUC reported for NFS ranged from 0.70 (0.62-0.78) to 0.83 (0.73-0.93). The accuracy of APRI was lower compared to FIB-4 and NFS in all seven studies, with AUC ranging from 0.52 to 0.73 (0.60-0.86). Four of four studies showed significant results ($P < .05$).[24-26,31] Here also, length of follow-up did not seem to influence the performance of any biomarker.

Of the studies that reported the threshold values used for prognosticating mortality, all used either or both the suggested high and low thresholds for FIB-4 and APRI. For NFS, thresholds of −0.9 and −1.836 were also studied in addition to the suggested thresholds. We again found sparse reporting of sensitivity and specificity. One

study found that at the high threshold, FIB-4, NFS and APRI showed sensitivity and specificity of 0.70 and 0.72, 0.69 and 0.76, and 0.55 and 0.89 respectively.[32]

4 | DISCUSSION

Non-invasive markers with comparable ability to prognosticate severe liver-related outcomes may be valuable tools for stratifying patients with higher risk of complication, in place of a liver biopsy. In this systematic review, we aimed to summarize the evidence on the prognostic performance of three multimarker models in identifying those at risk of developing worsening of NAFLD-related outcomes. We found that FIB-4, NFS and APRI have limited performance in predicting changes in fibrosis, as evaluated by future biopsies, but consistently demonstrated the ability to predict liver-related morbidity and mortality, with a level of performance that met or exceeded that of a liver biopsy.

4.1 | Strengths and limitations

While many studies have synthesized data on the diagnostic accuracy of non-invasive NAFLD markers, to our knowledge, this is the first systematic review conducted on the prognostic context of use. In collaboration with a search specialist, we developed a highly sensitive search strategy, including abstracts, to minimize bias that may arise from selective inclusion. For robust evaluation of bias in individual studies, we used a new risk of bias tool developed specifically for systematic reviews of prognostic accuracy.[16] All screening phases, data extraction and quality assessment were independently conducted by two experienced methodologists.

Our work comes with limitations, some inherent to the nature of prognostic research. Several studies had a relatively short follow-up period. This can be problematic for assessing outcomes of a chronic condition such as NAFLD, where patients have a median survival period of >10 years.[33,34] The results should be interpreted with caution, given the limited and heterogeneous follow-up periods, which ranged from one to 20 years. The variability in study

TABLE 2 Accuracy of biomarkers FIB-4, NFS and APRI in prognosticating change in fibrosis stage, liver-related events and mortality among NAFLD patients

Author	Target event	No. of cases (%)[a]	Time horizon (years)	AUC/C-index FIB-4	NFS	APRI
Fibrosis						
Vilar-Gomez (2017)	Fibrosis progression[a]	45 (17)	1	0.65 (0.54-0.76)	0.69 (0.58-0.79)	0.65 (0.53-0.73)
Chalasani (2018)	Fibrosis progression[a]	NA	1.4	0.68 (0.60-0.76)	0.65 (0.56-0.73)	0.72 (0.65-0.80)
Siddiqui (2019)	Fibrosis progression[a]	92 (32)	2.6	0.73 (0.67-0.79)	0.66 (0.59-0.73)	0.70 (0.63-0.77)
McPherson (2015)	Progression to fibrosis stage ≥ 3	46 (43)	6.6	NA	0.83 (0.74-0.92)[*]	0.72 (0.62-0.82)[*]
Siddiqui (2019)	Progression to fibrosis stage ≥ 3	35 (16)	2.6	0.81 (0.73-0.89)	0.80 (0.71-0.88)	0.82 (0.74-0.89)
Vilar-Gomez (2017)	Fibrosis regression[b]	51 (20)	1	0.57 (0.51-0.68)	0.63 (0.58-0.75)	0.59 (0.52-0.70)
Liver-related events						
Ioannou (2019)	HCC[c]	407 (6)	3.7	0.71	NA	NA
Peleg (2018)	Liver-related events[d]	86 (56)	1.9	0.89	0.92	0.73
Angulo (2013)	Liver-related events[e]	60 (19)	8.7	0.86 (0.80-0.92)[*]	0.81 (0.76-0.87)[*]	0.80 (0.73-0.86)[*]
Onnerhag (2019)	Liver-related events[f]	20 (14)	17.7	0.81 (0.69-0.93)[*]	0.77 (0.64-0.89)[*]	0.82 (0.72-0.92)[*]
Hagstrom (2019)	Severe liver disease[g]	76 (12)	19.9	0.72	0.72	0.69
Sebastiani (2015)	Clinical outcomes[h]	25 (17)	5	0.79 (0.69-0.91)	0.89 (0.83-0.95)	0.89 (0.82-0.96)
Mortality						
Boursier (2016)	Liver-related mortality	17 (5)	6.4	0.78 (0.66-0.88)[*]	NA	0.69 (0.49-0.84)[*]
Peleg (2018)	All-cause mortality	19 (12)	1.9	0.78	0.80	0.63
Boursier (2016)	All-cause mortality	83 (23)	6.4	0.70 (0.64-0.75)	NA	0.54 (0.46-0.61)
Xun (2014)	All-cause mortality	12 (7)	6.6	0.81 (0.70-0.91)[**]	0.83 (0.73-0.93)[**]	0.73 (0.60-0.86)[**]
Angulo (2013)	All-cause mortality[i]	41 (13)	8.7	0.67 (0.58-0.76)[**]	0.70 (0.62-0.78)[*]	0.63 (0.53-0.72)[**]
Treeprasertsuk (2013)	All-cause mortality	39 (13)	11.9	NA	0.70	NA
Onnerhag (2019)	All-cause mortality	85 (59)	17.7	0.82 (0.75-0.90)[*]	0.82 (0.74-0.90)[*]	0.59 (0.50-0.68)
Hagstrom (2019)	All-cause mortality	214 (33)	19.9	0.72	0.72	0.52

Note: Cumulative incidence: number of new cases/number of persons at start of the observation period.

[a]Increase of at least 1 point in fibrosis score.

[b]Decrease of at least 1 point in fibrosis score.

[c]Hepatocellular carcinoma, defined as ICD-9 code 155.0 and ICD-10 code C22.0.

[d]Ascites, esophageal varices, hepatic encephalopathy, liver transplantation, TIPS or hospitalizations.

[e]Ascites, gastroesophageal varices/bleeding, portosystemic encephalopathy, spontaneous bacterial peritonitis, hepatocellular cancer, hepatopulmonary syndrome, or hepatorenal syndrome.

[f]Ascites, encephalopathy, variceal bleeding, or hepatocellular carcinoma.

[g]Cirrhosis, decompensated liver disease, liver failure, or hepatocellular carcinoma.

[h]Death, liver transplantation and end-stage hepatic complications defined as hepatocellular carcinoma, ascites, spontaneous bacterial peritonitis, hepatic encephalopathy, de novo varices or significant worsening of varices.

[i]Including liver transplant.

[*]$P < .001$.

[**]$P < .05$.

designs prohibited meta-analysis to produce summary estimates of performance.

In the scheme of disease management, risk stratification may be most beneficial in a primary care setting, in which the purpose is to identify patients who require expedited referral to tertiary care centres. All identified studies evaluated the markers prognostic performance in a secondary or tertiary care setting. Thus, data from these studies cannot necessarily be extrapolated to a primary care setting.

Furthermore, we observed that very few studies reported data on both threshold values and corresponding sensitivity and specificity, which are more informative and clinically relevant than the AUC alone. Sparse reporting may be attributed to the relatively new and therefore less established nature of prognostic accuracy studies in general, in comparison to diagnostic accuracy studies. Given the increased volume of prognostic accuracy research, reporting guidelines and quality assessment tools specific for this area of research should be further developed.

4.2 | In the context of current evidence

A 2015 editorial illustrated the prognostic value of histological features of NAFLD, in the form of a hierarchical model.[34] This model ranked fibrosis as the most important histological lesion associated with long-term outcomes in NAFLD, and many studies support biopsy-confirmed fibrosis to be a major prognostic marker for mortality.[35,36] However, growing literature highlights the limitations of a liver biopsy,[11] particularly for detection of fibrosis.[37] Aside from the risk of complications and invasive nature, sampling variability is a big concern. In a study by Ratziu et al where two biopsy samples were compared, fibrosis stage was different in 41% of patients.[38] This may not be surprising, as only 1/50 000 of a whole liver tissue is sampled during a biopsy.[39] Even for NASH, histological lesions are unevenly distributed throughout the liver tissue. Further problems with pathological diagnosis arise with inter- and intra-observer variability. Therefore, evaluating test accuracy with an imperfect reference standard such as a liver biopsy poses the risk of underestimating NASH and fibrosis severity.

While histological fibrosis predicts disease progression, prognostication of NAFLD-related events using non-invasive markers is an appealing alternative, especially if performance of these markers approximates or equals that of histology-confirmed fibrosis. In comparing the performance of non-invasive methods to histological fibrosis (F3-F4) in prognosticating liver-related events, APRI and FIB-4 had higher AUC compared to a biopsy, and the overall percent of accurate prognosis was higher for all three multimarker models (models had 84%-86% accuracy compared to 76% with a liver biopsy).[21] The AUC found in this study were consistent with others identified in this systematic review.

This direct comparison illustrated the ability of non-invasive markers to risk stratify patients with comparable, or even better performance than a liver biopsy. Another study supported this finding for the ELF test.[40] However, studies evaluating head-to-head comparisons of non-invasive markers and a liver biopsy are limited, and future studies should aim to validate these findings and build a stronger evidence-base for non-invasive tests, particularly for the simple multimarker models that contain components readily evaluated in routine laboratories.

In addition to FIB-4, NFS and APRI, other NAFLD markers have been studied for their prognostic ability. The ELF test is recognized by guidelines as a diagnostic marker for liver fibrosis. For predicting progression to cirrhosis and liver related events, the AUC for ELF was 0.79 and 0.68, respectively, out-performing histological assessment for both outcomes.[40] Vibration-controlled transient elastography (VCTE), a imaging technique validated for liver fibrosis, had an AUC of 0.73 (0.66-0.78) for all-cause mortality, significantly outperforming APRI (P = .001) but not FIB-4.[26] Liver stiffness measurement, by transient elastography (FibroScan) had an AUC of 0.86 (0.82-0.95) in prognosticating liver-related mortality in one study,[26] and an AUC of 0.911 (0.82-0.99) in prognosticating liver-related events.[41] Fibroscan significantly outperformed APRI for predicting all-cause mortality. FibroTest, another marker for determining stages of NAFLD-related

fibrosis, had an AUC of 0.94 (0.91-0.98) in prognostication of liver-related death.[19] The same study conducted a post hoc analysis comparing FibroTest and FIB-4 and found no significant difference in performance (P = .32). In this study, FIB-4 had an AUC of 0.87 (0.74-0.99). Longitudinal assessment of magnetic resonance elastography (MRE) showed prognostic accuracy of 0.62 (0.46-0.78) for predicting fibrosis improvement and magnetic resonance imaging (MRI-PDFF) had an AUC of 0.70 (0.57-0.83) for predicting steatosis reduction.[42] While some of these markers show promising results, more studies are needed to validate the findings.

5 | IMPLICATIONS FOR CURRENT PRACTICE

In clinical practice, FIB-4 and NFS can be used in regular intervals to detect disease progression, offering a less invasive, and perhaps a more accurate alternative to a biopsy. The annual change of NFS in patients who died was twofold that of survivors and, for fibrosis progression, fourfold higher in progressors than in those who were stable.[27] Another study found that FIB-4 and NFS were significantly higher among fibrosis progressors compared to non-progressors, despite no significant difference in histological grading.[28] Patients who underwent serial measurements of FIB-4 within 5 years and had high-risk in both occurrences had significantly increased risk of severe liver disease with an adjusted hazard ratio of 17.04 (11.67-24.88), and an accuracy of 98%.[43]

The costs and time invested into drug development has become increasingly exhaustive.[44] Given the volume of ongoing clinical trials for the treatment of NASH and fibrosis, and the understood complexities and required resources, prognostic markers can be an integral measure for expediting clinical trials. A marker linked to a clinical trial endpoint can improve efficiency for late stage clinical trials by identifying patients more likely to develop the outcome, ultimately reducing the number of participants recruited to a study.[45] For clinical trials targeting patients with cirrhosis, long-term events that characterize clinical decompensation (ascites, encephalopathy, HCC, variceal hemorrhage) are of interest.[46] We observed that all three markers showed consistently good prognostic performance for events indicating clinical decompensation.

In conclusion, this systematic review shows that FIB-4, NFS and APRI can risk stratify patients for liver-related morbidity and mortality, with comparable performance to a liver biopsy. If confirmed in future comparative studies with sufficient length of follow-up, the strong prognostic performance of these multimarker models could position them at the cornerstone for risk stratification and risk management among NAFLD patients.

ORCID

Jenny Lee https://orcid.org/0000-0003-4024-0933
Yasaman Vali https://orcid.org/0000-0001-7002-118X
Quentin M. Anstee https://orcid.org/0000-0002-9518-0088

REFERENCES

1. Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of non-alcoholic fatty liver disease: practice Guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. *Hepatology*. 2012;55(6):2005-2023.
2. Wong RJ, Cheung R, Ahmed A Nonalcoholic steatohepatitis is the most rapidly growing indication for liver transplantation in patients with hepatocellular carcinoma in the U.S. *Hepatology*. 2014;59(6):2188-2195.
3. Younossi Z, Anstee QM, Marietti M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol*. 2018;15(1):11-20.
4. Pais R, Barritt AST, Calmus Y, et al. NAFLD and liver transplantation: current burden and expected challenges. *J Hepatol*. 2016;65(6):1245-1257.
5. Anstee QM, Reeves HL, Kotsiliti E, Govaere O, Heikenwalder M. From NASH to HCC: current concepts and future challenges. *Nat Rev Gastroenterol Hepatol*. 2019;16(7):411-428.
6. Powell EE, Cooksley WG, Hanson R, Searle J, Halliday JW, Powell LW The natural history of nonalcoholic steatohepatitis: a follow-up study of forty-two patients for up to 21 years. *Hepatology*. 1990;11(1):74-80.
7. Fassio E, Alvarez E, Dominguez N, Landeira G, Longo C. Natural history of nonalcoholic steatohepatitis: a longitudinal study of repeat liver biopsies. *Hepatology*. 2004;40(4):820-826.
8. Adams LA, Lymp JF, St Sauver J, et al. The natural history of non-alcoholic fatty liver disease: a population-based cohort study. *Gastroenterology*. 2005;129(1):113-121.
9. Angulo P, Kleiner DE, Dam-Larsen S, et al. Liver fibrosis, but no other histologic features, is associated with long-term outcomes of patients with nonalcoholic fatty liver disease. *Gastroenterology*. 2015;149(2):389-397.e10.
10. Williams CD, Stengel J, Asike MI, et al. Prevalence of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis among a largely middle-aged population utilizing ultrasound and liver biopsy: a prospective study. *Gastroenterology*. 2011;140(1):124-131.
11. Lee DH Noninvasive evaluation of nonalcoholic fatty liver disease. *Endocrinol Metab*. 2020;35(2):243-259.
12. European Association for the Study of the L, European Association for the Study of D, European Association for the Study of O. EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. *J Hepatol*. 2016;64(6):1388-1402.
13. Siddiqui MS, Yamada G,Vuppalanchi R, et al. Diagnostic accuracy of noninvasive fibrosis models to detect change in fibrosis stage. *Clin Gastroenterol Hepatol*. 4:4.
14. Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology*. 2005;41(6):1313-1321.
15. Bedossa P, Arola J, Susan D, 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.
16. Lee J, Vali Y, Zafarmand M, Bossuyt P. Quality Assessment of Prognostic Accuracy Studies (QUAPAS): an extension of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool for systematic reviews of prognostic test accuracy studies Abstracts of the 26th Cochrane Colloquium, Santiago, Chile. *Cochrane Database Syst Rev* 2020; (1 Suppl 1). https://doi.org/10.1002/14651858.CD201901
17. 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(8):529-536.
18. Hagström H, Talbäck M, Andreasson A, Walldius G, Hammar N Ability of noninvasive scoring systems to identify individuals in the population at risk for severe liver disease. *Gastroenterology*. 2020;158(1):200-214.
19. Munteanu M, Pais R, Peta V, et al. Long-term prognostic value of the FibroTest in patients with non-alcoholic fatty liver disease, compared to chronic hepatitis C, B, and alcoholic liver disease. *Aliment Pharmacol Ther*. 2018;48(10):1117-1127.
20. Chalasani N, Abdelmalek MF, Loomba R,. et al. Relationship between three commonly used non-invasive fibrosis biomarkers and improvement in fibrosis stage in patients with non-alcoholic steatohepatitis. *Liver Int*. 39(5), 924-932.
21. Sebastiani G, Alshaalan R, Wong P, et al. Prognostic value of non-invasive fibrosis and steatosis tools, hepatic venous pressure gradient (HVPG) and histology in nonalcoholic steatohepatitis. *PLoS One*. 2015;10(6):e0128774.
22. Vilar-Gomez E, Calzadilla-Bertot L, Friedman SLet al. Serum biomarkers can predict a change in liver fibrosis 1 year after lifestyle intervention for biopsy-proven NASH. *Liver Int*. 37(12), 1887-1896.
23. Ioannou GN, Green P, Kerr KF, Berry K Models estimating risk of hepatocellular carcinoma in patients with alcohol or NAFLD-related cirrhosis for risk stratification. *J Hepatol*. 27:27.
24. Xun YH, Guo JC, Lou GQ et al. Non-alcoholic fatty liver disease (NAFLD) fibrosis score predicts 6.6-year overall mortality of Chinese patients with NAFLD. *Clin Exp Pharmacol Physiol*. 41(9), 643-649.
25. Angulo P, Bugianesi E, Bjornsson ES, et al. Simple noninvasive systems predict long-term outcomes of patients with nonalcoholic fatty liver disease. *Gastroenterology*. 145(4), 782-789.e4.
26. Boursier J, Vergniol J, Guillet A, et al. Diagnostic accuracy and prognostic significance of blood fibrosis tests and liver stiffness measurement by FibroScan in non-alcoholic fatty liver disease. *J Hepatol*. 65(3), 570-578.
27. Treeprasertsuk S, Bjornsson E, Enders F, Suwanwalaikorn S, Lindor KD. NAFLD fibrosis score: a prognostic predictor for mortality and liver complications among NAFLD patients. *World J Gastroenterol*. 19(8), 1219-1229.
28. McPherson S, Hardy T, Henderson E, Burt AD, Day CP, Anstee QM Evidence of NAFLD progression from steatosis to fibrosing-steatohepatitis using paired biopsies: implications for prognosis and clinical management. *J Hepatol*. 2015;62(5):1148-1155.
29. McPherson S, Stewart SF, Henderson E, Burt AD, Day CP Simple non-invasive fibrosis scoring systems can reliably exclude advanced fibrosis in patients with non-alcoholic fatty liver disease. *Gut*. 2010;59(9):1265-1269.
30. Hagstrom H, Nasr P, Ekstedt M, Stal P, Hultcrantz R, Kechagias S. Accuracy of noninvasive scoring systems in assessing risk of death and liver-related endpoints in patients with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol*. 17(6), 1148-1156.e4.
31. Onnerhag K, Hartman H, Nilsson PM, Lindgren S Non-invasive fibrosis scoring systems can predict future metabolic complications

and overall mortality in non-alcoholic fatty liver disease (NAFLD). *Scand J Gastroenterol*. 54(3), 328-334.

32. Peleg N, Sneh Arbib O, Issachar A, Cohen-Naftaly M, Braun M, Shlomai A Noninvasive scoring systems predict hepatic and extra-hepatic cancers in patients with nonalcoholic fatty liver disease. *PLoS One*. 2018;13(8):e0202393.

33. Calzadilla Bertot L, Adams LA The natural course of non-alcoholic fatty liver disease. *Int J Mol Sci*. 2016;17(5):774.

34. Loomba R, Chalasani N. The hierarchical model of NAFLD: prognostic significance of histologic features in NASH. *Gastroenterology*. 2015;149(2):278-281.

35. Angulo P, Kleiner DE, Dam-Larsen S, et al. Liver fibrosis, but no other histologic features, is associated with long-term outcomes of patients with nonalcoholic fatty liver disease. *Gastroenterology*. 2015;149(2):389-397.e10.

36. Taylor RS, Taylor RJ, Bayliss S, et al. Association between fibrosis stage and outcomes of patients with nonalcoholic fatty liver disease: a systematic review and meta-analysis. *Gastroenterology*. 2020;158(6):1611-1625.e12.

37. Sumida Y, Nakajima A, Itoh Y Limitations of liver biopsy and non-invasive diagnostic tests for the diagnosis of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. *World J Gastroenterol*. 2014;20(2):475-485.

38. Ratziu V, Charlotte F, Heurtier A, et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology*. 2005;128(7):1898-1906.

39. Goldstein NS, Hastah F, Galan MV, Gordon SC. Fibrosis heterogeneity in nonalcoholic steatohepatitis and hepatitis C virus needle core biopsy specimens. *Am J Clin Pathol*. 2005;123(3):382-387.

40. Sanyal AJ, Harrison SA, Ratziu V, et al. The natural history of advanced fibrosis due to nonalcoholic steatohepatitis: data from the simtuzumab trials. *Hepatology*. 16:16.

41. Shili-Masmoudi S, Wong GL-H, Hiriart J-B, et al. Liver stiffness measurement predicts long-term survival and complications in non-alcoholic fatty liver disease. *Liver Int*. 2020;40(3):581-589.

42. Jayakumar S, Middleton MS, Lawitz EJ, et al. Longitudinal correlations between MRE, MRI-PDFF, and liver histology in patients with non-alcoholic steatohepatitis: analysis of data from a phase II trial of selonsertib. *J Hepatol*. 70(1), 133-141.

43. Hagström H, Talbäck M, Andreasson A, Walldius G, Hammar N Repeated FIB-4 measurements can help identify individuals at risk of severe liver disease. *J Hepatol*. 2020.

44. Kola I, Landis J Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discovery*. 2004;3(8):711-716.

45. Bakhtiar R. Biomarkers in drug discovery and development. *J Pharmacol Toxicol Methods*. 2008;57(2):85-91.

46. Sanyal AJ, Brunt EM, Kleiner DE, et al. Endpoints and clinical trial design for nonalcoholic steatohepatitis. *Hepatology*. 2011;54(1):344-353.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.