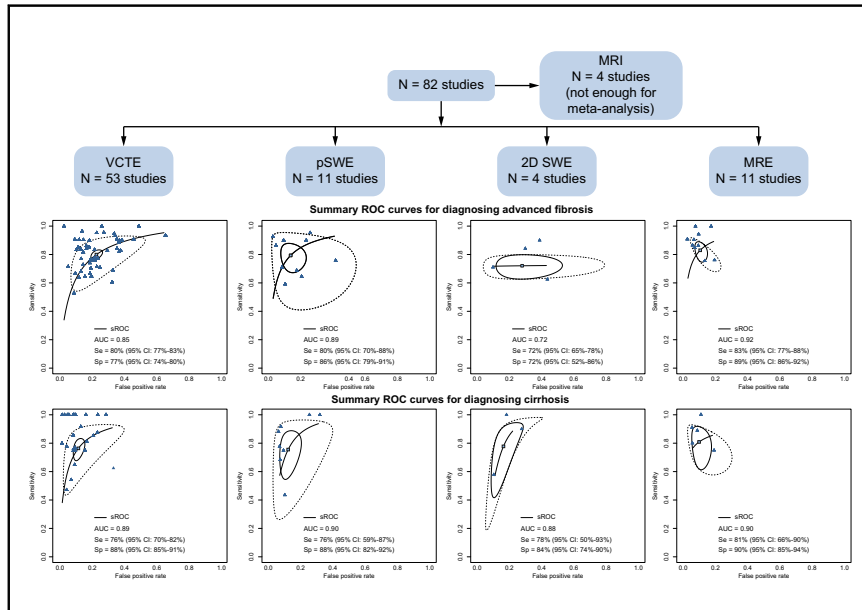# Diagnostic accuracy of elastography and magnetic resonance imaging in patients with NAFLD: A systematic review and meta-analysis

Graphical abstract

Highlights

- This is the largest systematic review of imaging/elastography biomarkers in NAFLD.

- Meta-analysis of 1 MR elastography and 3 ultrasound techniques.

- Elastography may help in fibrosis evaluation in those with NAFLD and valid readings.

- Clinical utility of these tests cannot be assessed fully as intention-to-diagnose analyses and validation of pre-specified cut-offs are lacking.

Authors

Emmanuel Anandraj Selvaraj, Ferenc Emil Mózes, Arjun Narayan Ajmer Jayaswal, …, Stephen A. Harrison, Patrick M. Bossuyt, Michael Pavlides

Correspondence

michael.pavlides@cardiov.ox.ac.uk
(M. Pavlides).

Lay summary

Non-invasive tests that measure liver stiffness or use magnetic resonance imaging (MRI) have been suggested as alternatives to liver biopsy for assessing the severity of liver scarring (fibrosis) and fatty inflammation (steatohepatitis) in patients with non-alcoholic fatty liver disease (NAFLD). In this study, we summarise the results of previously published studies on how accurately these non-invasive tests can diagnose liver fibrosis and inflammation, using liver biopsy as the reference. We found that some techniques that measure liver stiffness had a good performance for the diagnosis of severe liver scarring.

Diagnostic accuracy of elastography and magnetic resonance imaging in patients with NAFLD: A systematic review and meta-analysis

Emmanuel Anandraj Selvaraj[1,2,3,†], Ferenc Emil Mózes[1,†], Arjun Narayan Ajmer Jayaswal[1,†], Mohammad Hadi Zafarmand[4], Yasaman Vali[4], Jenny A. Lee[4], Christina Kim Levick[1], Liam Arnold Joseph Young[1], Naaventhan Palaniyappan[5], Chang-Hai Liu[6,7], Guruprasad Padur Aithal[5], Manuel Romero-Gómez[6], M. Julia Brosnan[8], Theresa A. Tuthill[8], Quentin M. Anstee[9,10], Stefan Neubauer[1], Stephen A. Harrison[1], Patrick M. Bossuyt[4], Michael Pavlides [1,2,3,*], on behalf of the LITMUS Investigators[#]

[1]Oxford Centre for Clinical Magnetic Resonance Research, Radcliffe Department of Medicine, University of Oxford, Oxford, UK; [2]Translational Gastroenterology Unit, Nuffield Department of Medicine, University of Oxford, Oxford, UK; [3]NIHR Oxford Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust and the University of Oxford, Oxford, UK; [4]Department of Epidemiology and Data Science, Amsterdam UMC, University of Amsterdam, The Netherlands; [5]NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, Nottingham, UK; [6]UCM Digestive Diseases, Virgen del Rocio University Hospital, Institute of Biomedicine of Seville, University of Seville, Sevilla, Spain; [7]Center for Infectious Diseases, West China Hospital of Sichuan University; Division of Infectious Diseases, State Key Laboratory of Biotherapy and Center of Infectious Disease, West China Hospital, Sichuan University, Chengdu, China; [8]Internal Medicine Research Unit, Pfizer Inc, Cambridge, MA, USA; [9]Liver Research Group, Translational & Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK; [10]NIHR Newcastle Biomedical Research Centre, Newcastle upon Tyne Hospitals, NHS Foundation Trust, Newcastle upon Tyne, UK

Background and Aims: Vibration-controlled transient elastography (VCTE), point shear wave elastography (pSWE), 2-dimensional shear wave elastography (2DSWE), magnetic resonance elastography (MRE), and magnetic resonance imaging (MRI) have been proposed as non-invasive tests for patients with non-alcoholic fatty liver disease (NAFLD). This study evaluated their diagnostic accuracy for liver fibrosis and non-alcoholic steatohepatitis (NASH).

Methods: PubMED/MEDLINE, EMBASE and the Cochrane Library were searched for studies examining the diagnostic accuracy of these index tests, against histology as the reference standard, in adult patients with NAFLD. Two authors independently screened and assessed methodological quality of studies and extracted data. Summary estimates of sensitivity, specificity and area under the curve (sAUC) were calculated for fibrosis stages and NASH, using a random effects bivariate logit-normal model.

Results: We included 82 studies (14,609 patients). Meta-analysis for diagnosing fibrosis stages was possible in 53 VCTE, 11 MRE, 12 pSWE and 4 2DSWE studies, and for diagnosing NASH in 4 MRE studies. sAUC for diagnosis of significant fibrosis were: 0.83 for VCTE, 0.91 for MRE, 0.86 for pSWE and 0.75 for 2DSWE. sAUC for diagnosis of advanced fibrosis were: 0.85 for VCTE, 0.92 for MRE, 0.89 for pSWE and 0.72 for 2DSWE. sAUC for diagnosis of cirrhosis were: 0.89 for VCTE, 0.90 for MRE, 0.90 for pSWE and 0.88 for 2DSWE. MRE had sAUC of 0.83 for diagnosis of NASH. Three (4%) studies reported intention-to-diagnose analyses and 15 (18%) studies reported diagnostic accuracy against pre-specified cut-offs.

Conclusions: When elastography index tests are acquired successfully, they have acceptable diagnostic accuracy for advanced fibrosis and cirrhosis. The potential clinical impact of these index tests cannot be assessed fully as intention-to-diagnose analyses and validation of pre-specified thresholds are lacking.

Lay summary: Non-invasive tests that measure liver stiffness or use magnetic resonance imaging (MRI) have been suggested as alternatives to liver biopsy for assessing the severity of liver scarring (fibrosis) and fatty inflammation (steatohepatitis) in patients with non-alcoholic fatty liver disease (NAFLD). In this study, we summarise the results of previously published studies on how accurately these non-invasive tests can diagnose liver fibrosis and inflammation, using liver biopsy as the reference. We found that some techniques that measure liver stiffness had a good performance for the diagnosis of severe liver scarring.

Introduction

Non-alcoholic fatty liver disease (NAFLD) is becoming the most common cause of end-stage liver disease worldwide, and is

strongly associated with metabolic syndrome (obesity, type 2 diabetes mellitus, dyslipidaemia and hypertension).[1] NAFLD encompasses a spectrum of conditions ranging from simple steatosis to non-alcoholic steatohepatitis (NASH) with or without liver fibrosis and cirrhosis.[2]

At the population level, most patients with NAFLD have simple steatosis and will not progress to more advanced stages of the disease. Even in patients who are identified as high risk and undergo liver biopsies, only a minority of patients will develop progressive fibrosis.[3] However, those with advanced fibrosis have poorer long-term outcomes.[4–6] Identifying this subgroup of high risk patients is one of the key issues in clinical care and drug trials. In the absence of any approved drug treatment, those at higher risk could benefit from lifestyle interventions and follow-up in secondary care. Furthermore, identifying patients with cirrhosis is important in order to enter them into surveillance for oesophageal varices and hepatocellular carcinoma. In clinical trials, diagnosis of NASH and histological staging of fibrosis are also important as these parameters define eligibility criteria and endpoints.

Histological classification of fibrosis and NASH remains the standard of practice, but the increasing burden of NAFLD renders its use unrealistic and inefficient. Liver biopsy is invasive, expensive, exhibits sampling variability, and is unacceptable to patients as a means of long-term dynamic monitoring of liver fibrosis stages.[7] Therefore, there has been much recent interest in developing robust, accurate and cost-effective non-invasive biomarkers to replace liver biopsy in severity assessment and risk stratification of patients with NAFLD.

Several non-invasive tests have emerged as promising alternatives for staging liver fibrosis and diagnosing NASH. Elastography-based techniques include vibration-controlled transient elastography (VCTE), point shear wave elastography (pSWE), 2-dimensional shear wave elastography (2DSWE), and magnetic resonance elastography (MRE). Magnetic resonance imaging (MRI) techniques include Liver*MultiScan*™ (LMS) to measure iron-corrected T1 (cT1), diffusion-weighted imaging (DWI), and detection of metabolic and liver injury (deMILI). Despite extensive experience with some of these index tests, significant questions remain about how best to use them in practice. None of these technologies have undergone sufficient validation to be granted regulatory approval for use in the context of clinical trials, while there remains a lack of consensus on thresholds for disease risk stratification in relation to histology.

In order to better understand the evolution of these tests and to summarise the literature to date, the aim of this study was to conduct a systematic review and meta-analysis evaluating the diagnostic performances of elastography and MRI index tests for the assessment of liver fibrosis and NASH in patients with NAFLD.

Materials and methods

The protocol for this systematic review is available on PROSPERO: CRD42018116522. This study is being reported according to the PRISMA-DTA guidelines (Table S1).

Target conditions

Liver fibrosis and NASH were the target conditions. Liver fibrosis was defined according to the NASH Clinical Research Network (CRN) histological classification.[8] The diagnostic accuracy of index tests was assessed in the following dichotomised groups: F0 *vs.* F1-4, F0-1 *vs.* F2-4, F0-2 *vs.* F3-4, F0-3 *vs.* F4, and NASH *vs.* simple steatosis. For the purpose of this review, any definition of NASH was accepted.

Index tests

The following index tests were assessed in this review: VCTE (FibroScan®, Echosens, Paris, France), pSWE (Virtual Touch Quantification (VTQ); Siemens Healthineers, Erlangen, Germany), 2DSWE (Aixplorer®; SuperSonic Imagine, Aix-en-Provence, France), MRE (Resoundant, Rochester, USA), cT1 measured using LMS (Perspectum, Oxford, UK), DWI, and deMILI. Each technique is summarised in Table S2.

We defined the technical failure as either unsuccessful valid measurements of an index test, unreliable measurement according to pre-defined quality criteria or poor-quality image acquisition such that analysis of data was not possible.

Inclusion criteria

Studies in all languages reported in peer-reviewed journals or conference abstracts were included if they fulfilled the following criteria: i) reporting on adults (≥18 years) with biopsy-proven NAFLD (data available on at least 10 patients); ii) index test performed within 6 months of biopsy; iii) liver histology according to the NASH CRN scoring system was used as the reference standard; iv) discrete data for NAFLD population could be extracted from mixed liver disease study cohort; v) estimates of the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) or receiver operating characteristic curves (ROCs) for diagnosing fibrosis stages and distinguishing NASH from simple steatosis were reported.

Exclusion criteria

Studies were excluded if they: i) included patients with coexisting liver disease (*e.g.* NAFLD and viral hepatitis in the same patient), ii) addressed a different context of use; iii) used an alternative histological classification system (*e.g.* METAVIR); iv) reported on using a pSWE or 2DSWE test other than what we specified in the index text section above; v) had insufficient data to calculate diagnostic accuracy estimates. For studies with missing data or where diagnostic performance was not reported separately for patients with NAFLD in a mixed liver disease cohort of patients, the corresponding or senior author was contacted by email to request the relevant data or results. The study was excluded if no reply was received within 28 calendar days.

Literature search

A systematic web-based literature search of all publications in PubMed/MEDLINE, EMBASE and CENTRAL (Cochrane Library) was conducted in March 2018, January 2019 and July 2020 (see Table S3 for details of the search terms). Reference lists of related systematic reviews and included studies were searched manually to identify additional studies.

Study selection

Search results were imported into an online platform for systematic review management (Covidence, Veritas Health Innovation, Melbourne, Australia. www.covidence.org) and duplicates were removed automatically. Titles and abstracts were screened first to identify potentially relevant papers, which were then assessed in full for eligibility. At least 2 researchers

conducted the screening of titles, abstracts, and full papers independently. Disagreements were resolved by reaching consensus between the researchers and if this was not possible, then a senior member of the team adjudicated. If multiple reports of the same study were identified, the most comprehensive and suitable publication related to our study was selected based on reaching a consensus among reviewers.

Data extraction

Two researchers independently extracted data using a standardised data extraction sheet. Disagreements were resolved by consensus or, where not possible, by arbitration from a senior member of the review team. Data was collected on the study characteristics (country, affiliation, year of publication, type of study), patient characteristics (age, sex, ethnicity, BMI, presence of metabolic syndrome, laboratory parameters), details of index text, performance indices of index test (cut-off values, failure rates, sensitivity, specificity, PPV, NPV, AUROC), and quality of liver biopsy and histological fibrosis stages. Necessary data to calculate the number of true positives, false positives, true negatives and false negatives were extracted. If this was not reported, they were calculated from diagnostic test sensitivity, specificity, and prevalence provided in the study.

Methodological quality assessment

Risk of bias and concerns about the applicability of study findings to the review question were assessed by 2 reviewers, independent of one another, using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool.[9] Disagreements were resolved by consensus, if possible, and adjudicated by a third member of the review team otherwise.

Evaluation of diagnostic accuracy

Classification tables were extracted and re-constructed for the performance of the index test for each of the pre-defined target conditions. For dichotomous classifications, study-specific estimates of sensitivity, specificity, PPV, NPV, positive likelihood ratio and negative likelihood ratio and their 95% CIs were calculated. Minimum acceptable performance of diagnostic accuracy was defined as sensitivity and specificity of at least 80%.[10] Graphical descriptive analysis of the included studies was performed using forest plots.

A meta-analysis was conducted whenever ≥3 studies with sufficient information for generating classification tables were available, within the same index test and target condition. In index tests without sufficient studies to conduct meta-analysis a narrative synthesis was conducted. When more than 1 cut-off value was presented, the cut-off value closest to the median value of all studies in the same group was selected for inclusion in the meta-analysis. We used a bivariate logit-normal random effects model to estimate the mean sensitivity, mean specificity and the respective variances and covariance. Summary receiver operator characteristic curves (sROC) were generated with 95% confidence regions and 95% prediction regions. The 95% confidence region is based on the confidence interval around the summary point and indicates that, based on the available data, we would expect the 'real value' to be within that region 95% of the time. The prediction region around the summary point indicates the region where we would expect results from a new study in the future to lie and is therefore wider than the confidence region as it goes beyond the uncertainty in the available

data. 95% CIs corresponding to summary AUC (sAUC) values were estimated via 500 bootstrap iterations.

A linear mixed effects model was used for modeling the multiple thresholds data of individual studies reporting more than 2 cut-offs.[11,12] The multiple thresholds model is a multi-level random effects model that enables the calculation of summary sensitivities and specificities of different cut-offs, and the calculation of the PPV and NPV, given the prevalence of the target condition. Sensitivity and specificity were combined at every recommended cut-off to produce a multiple-threshold sROC curve. In addition, PPV and NPV were also obtained, and cut-offs required to achieve minimum acceptable criteria were determined.

We did not attempt to construct funnel plots as it is well known that statistical tests based on funnel plot asymmetry cannot discriminate between publication bias and other sources of asymmetry, *e.g.* the effect of including multiple thresholds, in systematic reviews of diagnostic test accuracy studies.[13]

The statistical software R with the mada[14] and diagmeta[15] packages (Version 3.6.1; R Foundation for Statistical Computing, Vienna, Austria) was used in all analyses. *Post hoc* covariate meta-regression was performed for VCTE studies to explore potential sources of heterogeneity. The time interval between biopsy and VCTE, probe type and origin of study were examined as potential covariates. Reitsma-models were built using the mada R package with and without these covariates for each fibrosis stage group and compared using the likelihood ratio test statistic.

Results

Search results

A total of 13,819 articles were identified and imported into Covidence from the electronic databases searches. After removing duplicates, we screened 6,123 articles. We found 574 articles for full-text review from the electronic searches and 2 from hand-search of reference list. We were able to include 82 studies (75 full-text reports and 7 conference abstracts) in the systematic review. After excluding 12 studies with insufficient data, 70 studies were included in the meta-analysis as shown in Fig. 1.

Study characteristics

The characteristics of the VCTE,[16–80] MRE,[39,49,62,81–89] pSWE,[19,24,25,42,47,50,84,90–94] 2DSWE,[25,47,95,96] and MRI[30,36,63,97] studies included in the systematic review are summarised in Table 1. There were 73 prospective studies and 9 retrospective studies. Eleven studies compared 2 index tests and 2 studies compared 3 index tests. There were 55 single-centre and 27 multi-centre studies. Studies were from Europe (38%), Asia (38%), North America (18%), South America (4%) and Australia (2%). All studies were performed in a hospital setting.

Study quality

The methodological quality of the studies assessed with the QUADAS-2 tool is summarised in Figs. S1-S5. There were only 2 studies with no risk of bias or applicability concerns.[68,83] Studies that did not report pre-defined cut-off values were judged as having high risk of bias in the index test domain of QUADAS-2. This included 80% of VCTE, 86% of MRE, 92% of pSWE, 100% of 2DSWE and 76% of MRI studies. The flow and timing domain was judged to have high risk or unclear risk of bias in 80% of VCTE,
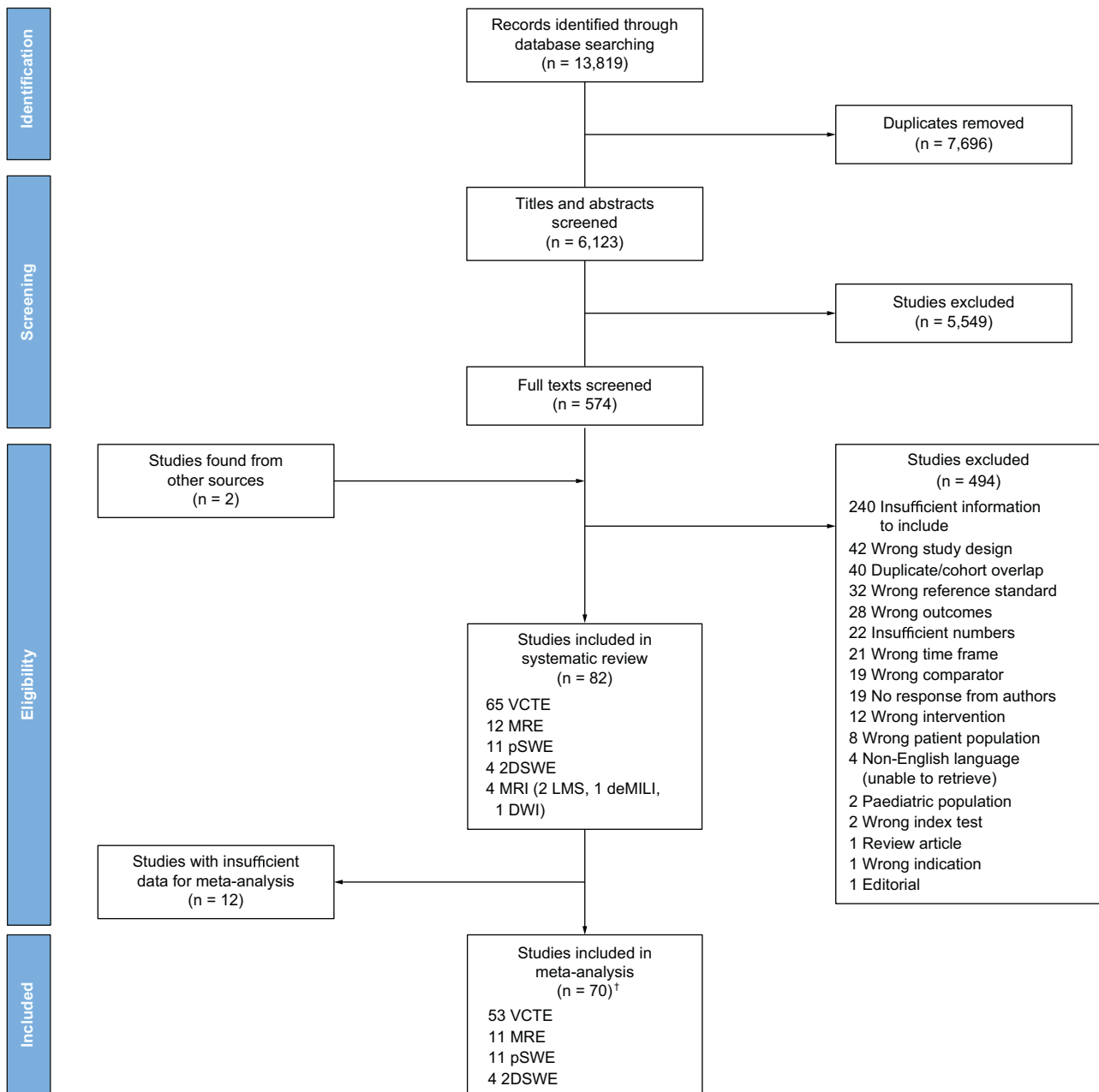
Fig. 1. PRISMA flow diagram of primary studies included in both the present systematic review and meta-analysis. 2DSWE, 2-dimensional shear wave elastography; deMILI, detection of metabolic and liver injury; DWI, diffusion-weighted imaging; LMS, Liver*MultiScan*™; MRE, magnetic resonance elastography; MRI, magnetic resonance imaging; pSWE, point shear wave elastography; VCTE, vibration-controlled transient elastography.

33% of MRE, 91% of pSWE, 100% of 2DSWE, and 75% of MRI studies as these studies either excluded technical failures from their final diagnostic performance analysis or did not report them.

Patient characteristics
In total, 14,609 patients with NAFLD were included in this review. There was a slight female preponderance (54%) with a mean or median age range of 35–63 years, mean or median BMI range of 27–48 kg/m², and 35% mean prevalence of diabetes in studies that reported this metric. The study populations included patients with biopsy-proven NAFLD (83%), biopsy-proven NAFLD

reported in a mixed liver disease aetiology cohort (9%), and patients from bariatric clinics or surgery (8%).

Index test characteristics
The range of technical failure of the index tests, when reported, were VCTE: 2–49%; MRE: 0–1%; pSWE: 0–43%; 2DSWE: 3–27%, and MRI: 5–18% (Table S4). Failure rates were not reported in 30% of VCTE, 39% of MRE, and 9% of pSWE studies. Cut-off values were pre-defined in only 20% of VCTE, 8% of MRE, 9% of pSWE, none of the 2DSWE and 25% of MRI studies (Table S5). Only 3 studies reported their intention-to-diagnose analyses.[60,73,76]

Table 1. Characteristics of studies included in the systematic review.

Ref.	Design	Population	Patients (n)	Age (years)	Male (%)	BMI (kg/m²)	T2DM (%)	HR (n; exp)	ITR (n;exp)
VCTE									
16	MC, P, CS	Chronic liver disease: Suspected NAFLD	25	–	–	–	–	1	–
17 ‡	SC, P, UC	Bariatric clinic	60	–	–	48 ± 7	–	1	–
18	MC, P, RCT	Biopsy-proven NASH	F0-2: 284	–	–	–	–	1	–
			F3-4: 1,323	–	–	–	–		
19	SC, P, CS	Suspected NAFLD	Overweight: 61	51 ± 13	51	28 ± 2	18	1	1
			Obese: 26	53 ± 10	58	36 ± 5	42		
20	SC, P, CS	Suspected NAFLD	88	46 ± 9	57	30 ± 5	19	1	–
21	MC, P, CS	Biopsy-proven NAFLD	452	56 ± 12	60	31 ± 5	47	1	1 (>500 examinations)
22	MC, P, CS	Biopsy-proven NAFLD	Training:625	56 ± 12	60	32 ± 6	51	>1	>1 (>500 examinations)
			Validation:313	57 ± 12	55	32 ± 6	52		
23	SC, P, CS	Suspected NAFLD	81	54 ± 10	26	33 ± 5	61	1	1
24	SC, P, CS	Chronic liver disease: Suspected NAFLD	M-probe: 48	55 ± 14	67	30 ± 5	48	1 (26 years)	>1 (>5,000 examinations)
			XL-probe: 49	55 ± 13	63	30 ± 5	43		
25	MC, P, CS	Biopsy-proven NAFLD	223	57 ± 12	59	32 ± 6	69	>1	>1 (>5,000 examinations)
26	SC, P, CS	Biopsy-proven NAFLD	Training: 101	50 ± 11	52	30 ± 4	53	1	>1
			Validation: 46	51 ± 13	61	29 ± 6	52		
27	MC, P, CC	Biopsy-proven NAFLD	57	50 ± 10	49	30 ± 5	–	1	
28 ‡	MC, P, CS	Suspected NAFLD	176	–	56	36	–	1	–
29	SC, P, CS	Chronic liver disease: Suspected NAFLD	25	46 ± 47	59	26 ± 4	–	1	–
30	MC, P, CS	Biopsy-proven NAFLD	47	–	50	34 ± 5	47	>1	>1
31	MC, P, CS	Suspected NAFLD	373	54 (19-77)†	55	34(9)†	52	2	>1
32	SC, P, CS	Biopsy-proven NASH	63	47 ± 8	62	–	–	1	1
33 ‡	SC, R, CS	Biopsy-proven NAFLD	238	–	–	31(5)†	–	–	–
34	SC, P, CS	Chronic liver disease: Suspected NAFLD	72	–	72	28	–	1	>1
35	SC, P, CS	Chronic liver disease: Suspected NAFLD	58	–	76	28	–	1	>1
36	MC, P, CS	Biopsy-proven NAFLD	126	51 ± 12	62	31 ± 5	37	–	–
37	SC, P, CS	Bariatric surgery	76	38 ± 10	21	45 ± 7	–	1	1
38 ‡	MC, R, UC	Biopsy-proven NAFLD	98	52 ± 11	43	37 ± 7	–	–	–
39	SC, P, CS	Biopsy-proven NAFLD	142	58 ± 15	57	28 ± 5	50	2	1
40	MC, P, CS	Biopsy-proven NAFLD	224	59 (17-85)†	46	28(17-44)†	55	2	>1 (>500 examinations)
41	SC, P, CS	Bariatric surgery	Training: 73	35 ± 8	32	41 ± 6	16	2	–
			Validation: 50	36 ± 9	26	40 ± 5	26		
42	SC, P, CS	Bariatric surgery and non-bariatric biopsy-proven NAFLD	Bariatric: 41	46 ± 10	32	47	–	1	–
			Non-bariatric: 45	55 ± 11	50	28 ± 4	–		
43	SC, P, CS	Biopsy-proven NAFLD	Non-cirrhotic: 120	39 ± 13	75	26 ± 4	17	2	–
			Cirrhotic: 85	53 ± 9	65	27 ± 4	48		
44	SC, P, CS	Suspected NAFLD in diabetes	94	–	–	–	100	1	>1 (>2,000 examinations, > 5 years)
45	SC, P, CS	Biopsy-proven NAFLD	126	–	–	–	–	1	>1
46	MC, P, CS	Suspected NAFLD	183	41 ± 14	61	28 ± 4	14	1	1
47	SC, P, CS	Biopsy-proven NAFLD	94	56 ± 13	44	27 ± 4	39	1	1 (>1,000 examinations)
48	SC, R, CS	Biopsy-proven NASH	184	45 ± 15	69	29†	38	1	–
49	SC, P, CS	Biopsy-proven NAFLD	130	51(41-62)†	41	30(26-33)†	42	2	2
50	SC, P, CS	Biopsy-proven NAFLD	100	–	–	–	–	1	2 (>200 examinations)
51	SC, R, CS	Suspected NAFLD	215	–	55	27(25-29)†	55	2	>1 (>50 examinations)
52	SC, P, CS	Biopsy-proven NASH	72	–	71	29†	–	1	–
53	SC, P, CS	Biopsy-proven NAFLD	131	50 ± 12	53	–	47	1	–
54 ‡	MC, P, CS	Biopsy-proven NAFLD	162	–	–	–	–	–	–
55	MC, P, CS	Chronic liver disease: Suspected NAFLD	75	–	–	–	56	2	9, (4 >500 examinations; 1 >200 examinations; 3 >100 examinations; and 1 >50 examinations)

Table 1. (*continued*)

Ref.	Design	Population	Patients (n)	Age (years)	Male (%)	BMI (kg/m^2)	T2DM (%)	HR (n; exp)	ITR (n;exp)
56	SC, P, CS	Bariatric surgery	100	43 ± 1	19	42 ± 1	15	1	>1
57	SC, P, CS	Bariatric surgery	Retrospective: 194	41 ± 1	22	44	18	1	>1
			Prospective: 123	40 ± 1	26	44 ± 1	22		
58	MC, P, CS	Suspected NAFLD	Training: 96	61(20-84)	43	28[†]	60	2	1 (>500 examinations)
			Validation: 103	63(23-90)	45	28[†]	65		
59	SC, P, CS	Biopsy-proven NAFLD	163	56 ± 14	49	27 ± 4	—	1	>1
60	MC, P, CS	Bariatric surgery	66	46 ± 12	32	45 ± 9	—	1	2 (>2,000 examinations)
61 [‡]	SC, P, CS	Biopsy-proven NAFLD	208	53 ± 12	67	—	30	—	—
62	SC, P, CS	Biopsy-proven NAFLD	97	51 ± 15	43	30 ± 5	30	1	1
63	SC, P, CS	Suspected NAFLD	71	53 ± 12	61	33(28-38)[†]	35	2	>1
64	SC, P, CS	Biopsy-proven NAFLD	146	44 ± 13	71	29 ± 4	14	1	1 (>100 examinations)
65	MC, R, CS	Biopsy-proven NAFLD	Training: 179	45 ± 13	68	29 ± 4	20	1	1 (>100 examinations)
			Validation: 142	44 ± 12	72	27 ± 3	16		
66	SC, P, CS	Biopsy-proven NAFLD	253	45 ± 13	70	29 ± 4	20	1	1 (>300 examinations)
67	MC, P, CS	Biopsy-proven NAFLD	761	51 ± 13	60	30 ± 5	55	1	1 (>300 examinations)
68	MC, P, CS	Biopsy-proven NAFLD	324	54 ± 13	44	—	46	1	1 (>300 examinations)
69	SC, P, CS	Chronic liver disease: Suspected NAFLD	105	45 ± 12	72	28 ± 4	—	1	—
70	SC, P, CS	Biopsy-proven NAFLD	171	57 ± 14	50	28 ± 5	—	1	—
71	MC, P, CS	Biopsy-proven NAFLD	101	—	—	—	—	3	>1, >300 examinations
72	SC, P, CS	Biopsy-proven NAFLD	249	58 ± 15	48	27 ± 4	57	1	1
73	MC, P, CS	Suspected NAFLD	140	—	—	—	—	2	—
74	SC, P, CS	Biopsy-proven NAFLD	120	50 ± 13	63	31(29-35)[†]	23	>1	2 (>500 examinations)
75	MC, P, CS	Biopsy-proven NAFLD	246	51 ± 11	55	28 ± 5	36	2	>1
76	MC, P, CS	Biopsy-proven NAFLD	193	52 ± 11	57	29 ± 5	51	2	>1 (>50 examinations)
77	MC, P, CS	Biopsy-proven NAFLD	496	—	—	—	—	2	>1 (>50 examinations)
78	MC, P, CS	Biopsy-proven NAFLD	97	52 ± 14	41	29 ± 4	—	2	1
79	MC, P, CS	Biopsy-proven NAFLD	292	45 ± 13	46	32 ± 7	11	>1	1
80	SC, P, CS	Chronic liver disease: Suspected NAFLD	13	—	77	31	—	—	—
MRE									
81	SC, R, CS	Biopsy-proven NAFLD	58	52	17	38	—	1 (6 years)	1 (4 years)
82	SC, P, CS	Biopsy-proven NAFLD	49	54 ± 13	14	32 ± 5	—	—	1 (15 years)
83	SC, P, CS	Biopsy-proven NAFLD	102	51 ± 14	59	32 ± 6	26	1	1 (≥6 months)
84	SC, P, CS	Biopsy-proven NAFLD	125	49 ± 15	46	32 ± 7	26	1	1 (≥6 months)
39	SC, P, CC	Biopsy-proven NAFLD	142	58 ± 15	57	28 ± 5	50	2	>1
85	SC, R, CS	Suspected NAFLD	142	53 ± 13	27	36 ± 7	28	>1	>2
86	SC, P, CS	Suspected NASH	47	51 ± 13	34	28 ± 6	—	1 (>15 years)	2 (>25 and >6 years)
49	SC, P, CS	Biopsy-proven NAFLD	130	51(41-62)[†]	41	30(26-33)[†]	42	2	1
87 [‡]	SC, P, CS	Biopsy-proven NAFLD	52	50 ± 13	52	32 ± 5	—	1	1 (≥6 months)
88	SC, P, CS	Biopsy-proven NAFLD	117	50 ± 13	44	32 ± 5	34	1	1 (≥6 months)
89	SC, P, CS	Biopsy-proven NAFLD	99	50 ± 14	44	32 ± 5	33	1	1 (≥6 months)
62	SC, P, CS	Suspected NAFLD	104	51 ± 15	43	30 ± 5	28	1	1 (≥6 months)
pSWE									
19	SC, P, CS	Suspected NAFLD	Overweight: 61	51 ± 13	51	28 ± 2	18	1	1
			Obese: 26	53 ± 10	58	36 ± 5	42		
24	SC, P, CS	Chronic liver disease: Suspected NAFLD	60	56 ± 13	67	30 ± 5	43	1 (26 years)	4 (9-11 years, >6 months ARFI, >100 examinations)
25	MC, P, CS	Biopsy-proven NAFLD	236	57 ± 12	59	32 ± 6	65	>1	6 (>2 years)
84	SC, P, CS	Biopsy-proven NAFLD	125	49 ± 15	46	32 ± 7	26	1	1
90	SC, P, CS	Biopsy-proven NAFLD	Simple steatosis: 21	47	40	29	—	1 (25 years)	—
			NASH: 43	51	47	30			
91	SC, P, CS	Biopsy-proven NAFLD	315	55	51	27[†]	38	1	>1

Table 1. *(continued)*

Ref.	Design	Population	Patients (n)	Age (years)	Male (%)	BMI (kg/m²)	T2DM (%)	HR (n; exp)	ITR (n;exp)
42	SC, P, CS	Bariatric surgery	Bariatric: 41	46 ± 10	32	47	–	1	>1
			Non-bariatric: 48	55 ± 11	50	28 ± 4	45		
47	SC, P, CS	Suspected NAFLD	83	56 ± 13	44	27 ± 4	–	1	2 (10 and 13 years)
92	SC, P, CS	Suspected and biopsy-proven NAFLD	51	–	–	–	–	>1	>1 (3–20 years)
93	SC, P, CS	Biopsy-proven NAFLD	135	–	38	–	–	1	5
94	SC, R, CS	Biopsy-proven NAFLD	67	35 ± 13	69	–	–	1	–
2DSWE									
25	MC, P, CS	Biopsy-proven NAFLD	232	57 ± 12	59	32 ± 6	66	1	>1
47	SC, P, CS	Suspected NAFLD	83	56 ± 13	44	27 ± 4	45		1 (12 years)
95	SC, R, CS	Suspected and biopsy-proven NAFLD	116	51 ± 12	47	31 ± 5	33	1 (30 years)	6
96	SC, P, CS	Biopsy-proven NAFLD	71	51 ± 16	65	29 ± 5	–	1	1 (>10 years ultrasound, >1 year SWE)
MRI									
30	MC, P, CS	Biopsy-proven NAFLD	50	54(18–73)†	56	34 ± 5	44	>1	–
36	MC, P, CS	Biopsy-proven NAFLD	126	51 ± 12	62	31 ± 5	37	1	2
97	SC, P, CS	Suspected NAFLD	59	54 ± 9	17	32	100	1 (28 years)	1 (10 years)
63	SC, P, CS	Suspected NAFLD	71	53 ± 12	60	33 (28-38)†	35	2	2

– Not reported or unable to derive.
CC, case-control; CS, cross-sectional; exp, experience; HR, histology readers; ITR, index test readers; MC, multi-centre; P, prospective; R, retrospective; SC, single-centre; T2DM, type 2 diabetes mellitus.
†Reported as median (range).
‡Abstracts.

Liver biopsy characteristics

Biopsy samples were evaluated by more than 1 pathologist in 29% of studies and a single pathologist in 62% of studies. It was not clear how the biopsies were reported in the remaining 9% of studies (Table 1). Consensus was sought between pathologists in 5% of studies. The size of the biopsy needle was reported in 43% of studies, length of biopsy specimen (or minimum acceptable quality criteria) in 67% of studies and number of portal tracts (or minimum acceptable quality criteria) in 39% of studies (Table S6).

Results of meta-analysis

Diagnosis of any fibrosis (F0 vs. F1-4)

The diagnostic accuracy in detecting any degree of fibrosis (≥F1) was investigated by the fewest studies (14 VCTE (n = 1,064), 6 MRE (n = 391), and 4 pSWE (n = 276); Table 2; Figs. S6-S8 for forest plots). These studies reported the poorest classification performance and no index test met the minimum acceptable performance for diagnostic accuracy. The respective sAUC, sensitivity and specificity for diagnosing stage ≥F1 were VCTE: 0.82, 78%, 72%; MRE: 0.87, 71%, 85%; and pSWE: 0.77, 64%, 76%. The summary point estimate of the mean with a 95% confidence region and 95% prediction region for each index test is shown in Fig. 2.

Diagnosis of significant fibrosis (F0-1 vs. F2-4)

The diagnostic accuracy in detecting significant fibrosis (≥F2) was investigated in 37 VCTE (n = 2,763), 6 MRE (n = 209), 9 pSWE (n = 805), and 4 2DSWE (n = 488) studies (Table 2; Figs. S9-S12 for forest plots). None of the index tests met the minimum acceptable performance for diagnostic accuracy. The respective sAUC, sensitivity and specificity for diagnosing stage ≥F2 were VCTE: 0.83, 80%, 73%; MRE: 0.91, 78%, 89%; pSWE: 0.86, 69%, 85%; and 2DSWE: 0.75, 71%, 67%. The summary point estimate of the mean, with a 95% confidence region and 95% prediction region for each index test is shown in Fig. 3.

Diagnosis of advanced fibrosis (F0-2 vs. F3-4)

The diagnostic accuracy in detecting advanced fibrosis (≥F3) was investigated by most studies (44 VCTE (n = 4,219), 10 MRE (n = 214), 11 pSWE (n = 1,209), and 4 2DSWE (n = 488); Table 2; Figs. S13-S16 for forest plots). The respective sAUC, sensitivity and specificity for diagnosing stage ≥F3 were VCTE: 0.85, 80%, 77%; MRE: 0.92, 83%, 89%; pSWE: 0.89, 80%, 86%; and 2DSWE: 0.72, 72%, 72%. MRE and pSWE met the minimum acceptable criteria for diagnostic accuracy. The summary point estimate of the mean with a 95% confidence region and 95% prediction region for each index test is shown in Fig. 4.

A multiple-threshold meta-analysis was performed in 6 primary studies (n = 1,278) reporting more than 2 cut-offs for VCTE. The sAUC was 0.85, and the Youden-index was maximised by an 8.7 kPa cut-off with 80% sensitivity and 76% specificity (Fig. S17). Predictive values for various cut-offs and prevalences are presented in Table S7. No cut-off met the minimum acceptance criteria of providing a sensitivity and specificity of 80%. However, a cut-off of 8.9 kPa was associated with 80% sensitivity and 77% specificity and a cut-off of 9.5 kPa was associated with 76% sensitivity and 80% specificity.

Diagnosis of cirrhosis (F0-3 vs. F4)

The diagnostic accuracy in detecting cirrhosis (F4) was investigated in 22 VCTE (n = 337), 5 MRE (n = 41), 8 pSWE (n = 759), and

Table 2. Summary diagnostic performance of VCTE, MRE, pSWE and 2DSWE for the detection of fibrosis stages in NAFLD, and diagnostic performance of MRE for the diagnosis of NASH.

	Studies, n (patients; n)	Prevalence, % (95% CI)	Cut-off range	sAUC (95%CI)	sSe, % (95% CI)	sSp, % (95% CI)
VCTE (kPa)						
F≥1	14 (1,064)	67 (23–94)	5.3–8.2	0.82 (0.78–0.85)	78 (73–82)	72 (65–79)
F≥2	37 (2,763)	45 (5–77)	3.8–10.2	0.83 (0.80–0.87)	80 (76–83)	73 (68–77)
F≥3	44 (4,219)	25 (5–54)	6.8–12.9	0.85 (0.83–0.87)	80 (77–83)	77 (74–80)
F=4	22 (337)	9 (3–31)	6.9–19.4	0.89 (0.84–0.93)	76 (70–82)	88 (85–91)
MRE (kPa)						
F≥1	6 (391)	60 (54–90)	2.50–3.14	0.87 (0.80–0.94)	71 (60–81)	85 (78–91)
F≥2	6 (209)	31 (25–54)	2.86–4.14	0.91 (0.80–0.97)	78 (67–85)	89 (83–94)
F≥3	10 (214)	19 (12–32)	2.99–4.80	0.92 (0.88–0.95)	**83 (77–88)**	**89 (86–92)**
F=4	5 (41)	8 (6–9)	3.35–6.70	0.90 (0.81–0.95)	**81 (66–90)**	**90 (85–94)**
NASH	4 (224)	69 (51–78)	2.53–3.26	0.83 (0.69–0.91)	65 (46–80)	83 (69–91)
pSWE (m/s)						
F≥1	4 (276)	73 (58–95)	1.11–1.81	0.77 (0.55–0.92)	64 (48–77)	76 (65–84)
F≥2	9 (805)	46 (17–73)	1.18–1.81	0.86 (0.78–0.90)	69 (59–77)	85 (80–88)
F≥3	11 (1,209)	30 (17–52)	1.34–4.24	0.89 (0.83–0.95)	**80 (70–88)**	**86 (82–92)**
F=4	8 (759)	17 (6–32)	1.36–2.54	0.90 (0.82–0.95)	76 (59–87)	88 (82–92)
2DSWE (kPa)						
F≥2	4 (488)	55 (26–71)	8.3–11.6	0.75 (0.58–0.87)	71 (56–83)	67 (43–84)
F≥3	4 (488)	36 (16–45)	9.3–13.1	0.72 (0.60–0.84)	72 (65–78)	72 (52–86)
F=4	3 (372)	15 (7–16)	14.4–15.7	0.88 (0.81–0.91)	78 (50–93)	84 (74–90)

Bold text indicates that the test met the minimum acceptable criteria of at least 80% sensitivity and specificity.
sAUC, summary area under the curve; sSe, summary sensitivity; sSp, summary specificity.

3 2DSWE (n = 372) studies (Table 2; Figs. S18-S21 for forest plots). The respective sAUC, sensitivity and specificity for diagnosing stage F4 were VCTE: 0.89, 76%, 88%; MRE: 0.90, 81%, 90%; pSWE: 0.90, 76%, 88%; and 2DSWE: 0.88, 78%, 84%. Only MRE met the minimum acceptable criteria for diagnostic accuracy. The summary point estimate of the mean with a 95% confidence region and 95% prediction region for each index test modality is shown in Fig. 5.

Diagnosis of steatohepatitis (NASH vs. simple steatosis)
There were 5 VCTE studies[30,39,46,49,62] and 1 pSWE study[90] that reported the diagnostic accuracy of liver stiffness in distinguishing NASH from simple steatosis. Data pooling for meta-analysis in the VCTE papers was not possible due to variability in reporting performance characteristics. The diagnostic accuracy in detecting NASH was investigated in 4 MRE (n = 224) studies. The sAUC, sensitivity and specificity were 0.83, 65% and 83%, respectively (Table 2, Fig. 6), and these did not meet the minimum acceptable criteria for diagnostic accuracy.

Narrative synthesis of MRI techniques
A narrative synthesis of the MRI results is included in the supplementary information.

Exploratory study of sources of heterogeneity in VCTE studies
Neither probe type nor study origin defined by the continent where the study was conducted were significant covariates of diagnostic performance for any of the fibrosis stages. Additionally, we found no significant difference in the diagnostic performance when comparing studies only allowing 3 months between VCTE and biopsy to all studies. Complete results of covariate testing, as well as sensitivity and subgroup analyses can be found in Tables S8-S11.

Discussion
There is an increasing clinical and research need to reduce reliance on liver biopsy to assess NAFLD disease severity given its increasing prevalence worldwide. In this study, we conducted a systematic review of 82 studies (14,609 patients) and meta-analysis of 70 studies (12,547 patients) to summarise the evidence for the diagnostic accuracy of 5 elastography and imaging modalities in the non-invasive diagnosis of liver fibrosis and NASH in adult patients with NAFLD.

We defined the minimum acceptable performance criteria of greater than 80% for both sensitivity and specificity as the benchmark for diagnostic accuracy tests in NAFLD. In those patients with successful measurements of liver stiffness, these criteria were met by MRE and pSWE for the diagnosis of advanced fibrosis and by MRE for the diagnosis of cirrhosis, with the caveat that the lower limit of the 95% CIs of summary sensitivities was <80%. Further validation of these tests is therefore needed before they can be confidently recommended as alternatives to liver biopsy.

Whilst the diagnostic performance for both MRE and pSWE were similar with AUC >0.90, the 95% confidence and prediction regions for MRE appear to be smaller than for pSWE, suggesting that there was less heterogeneity in the MRE studies.

MRE was the only modality with sufficient studies for meta-analysis for the diagnosis of NASH. Even though the diagnostic accuracy was good, this did not reach the pre-defined minimum acceptable criteria defined in our study.

Meta-analysis on MRI data was impossible due to the low number of primary studies, and, as a result, the performance of cT1 by LMS, deMILI and DWI could not be evaluated using the minimum acceptable criteria. We do note, however, that cT1 had typically high sensitivity and low specificity, deMILI had moderate sensitivity and specificity and DWI had poor to moderate sensitivitiy and specificity.

VCTE was the modality with most available data. Even though it did not meet the minimum acceptable criteria for any of the target conditions, these results should be interpreted with some caution as some studies did not use the XL probe or may not have used it according to the manufacturer's recommendations. We did however find that the probe used was not a significant factor of heterogeneity in the VCTE studies. Recent improvements of
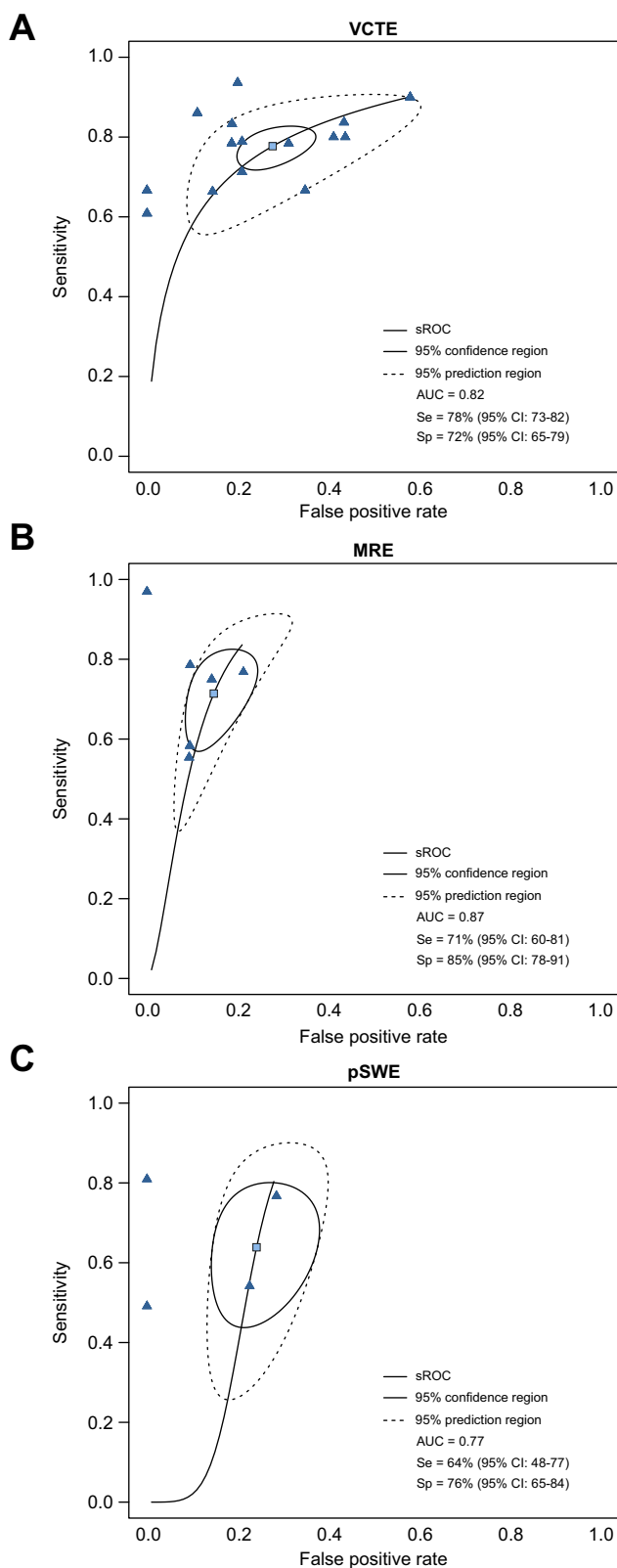
Fig. 2. sROC curves and test performance to detect any degree of fibrosis (F0 vs. F1-4). A bivariate logit-normal random effects model was used to estimate the mean sensitivity and mean specificity for liver stiffness measured using (A) VCTE, (B) MRE, and (C) pSWE. MRE, magnetic resonance elastography; pSWE, point shear wave elastography; sROC, summary receiver operating characteristic; VCTE, vibration-controlled transient elastography.

VCTE have been reported,[98,99] but these were not included in this study.

Several systematic reviews and meta-analyses (Table S12) have evaluated the diagnostic performance of different elastography and imaging modalities, but none so far have reported on the 5 main index tests as presented here. In particular, we found no previous meta-analyses for 2DSWE. The most recent pooled analysis of diagnostic accuracy by Liang *et al.* (12 studies, 910 patients),[100] examined the performance of MRE for staging liver fibrosis in NAFLD, and reported a better performance across all fibrosis stages with wider cut-off ranges, than the current and previous meta-analyses.[101,102] However, Liang *et al.*[100] included data from 4 mixed liver disease aetiology studies that we did not include as the studies did not use the NASH CRN histology scoring system, or discrete data for patients with NAFLD were not available. A unique feature of the present study is the analysis of the diagnostic performance of MRE in distinguishing NASH from simple steatosis as a recent meta-analysis only carried out a narrative synthesis of the data.[103]

Xiao *et al.* 2017 (64 studies, 13,046 patients with NAFLD)[102] comprehensively examined the diagnostic performance of VCTE and MRE alongside 4 serum tests for staging hepatic fibrosis against liver biopsy.[102] Whilst a large proportion of the included patients had NAFLD, their study data was more heterogeneous as 203 patients with viral hepatitis, 45 patients with other liver diseases, and 554 children were also included. Moreover, the time interval between index tests and liver biopsy was undefined, and studies using 3 other histological classification systems were also included.

The diagnostic accuracy of pSWE exclusively in patients with NAFLD was examined in a meta-analysis by Jiang *et al.* 2018 (9 studies, 982 patients),[104] which reported higher summary sensitivities in detecting advanced fibrosis and cirrhosis than our study, despite having similar sAUCs.[104] We, however, excluded 3 studies that were included by Jiang *et al.* due to the time interval between pSWE and biopsy being over 6 months, inclusion of a subset of paediatric population and duplication of data from the same study, respectively. The authors also reported a rather optimistically low rate of failed measurements (<1%) compared to us, most likely due to differences in the definition of a technical failure.

Most of the primary studies reported the experience of the index test operators or readers, however, data on intra- or inter-observer agreements were lacking. Similarly, the inter- and intra-observer variability in liver biopsy interpretation is well-known, particularly in classifying the intermediate stages of fibrosis. Despite this fact, the histology was read by a single pathologist in many studies. If liver histology is to be used as the reference standard, it would be preferable that at least 2 histopathologists review the liver biopsy specimen, preferably with consensus.

Our review has identified several areas where data in the literature are lacking. All but 2 of the studies we included in this systematic review and meta-analysis were scored as being at high risk of bias in at least 1 domain. This was mainly due to the fact that very few studies conducted validation of pre-defined cut-offs or intention-to-diagnose analyses. Without prospective validation of pre-defined cut-offs it is difficult to know how clinicians could apply these tests, while the lack of intention-to-diagnose analyses makes it difficult to fully evaluate the true impact of these tests in clinical practice. The true applicability is also difficult to evaluate as many studies did not report success
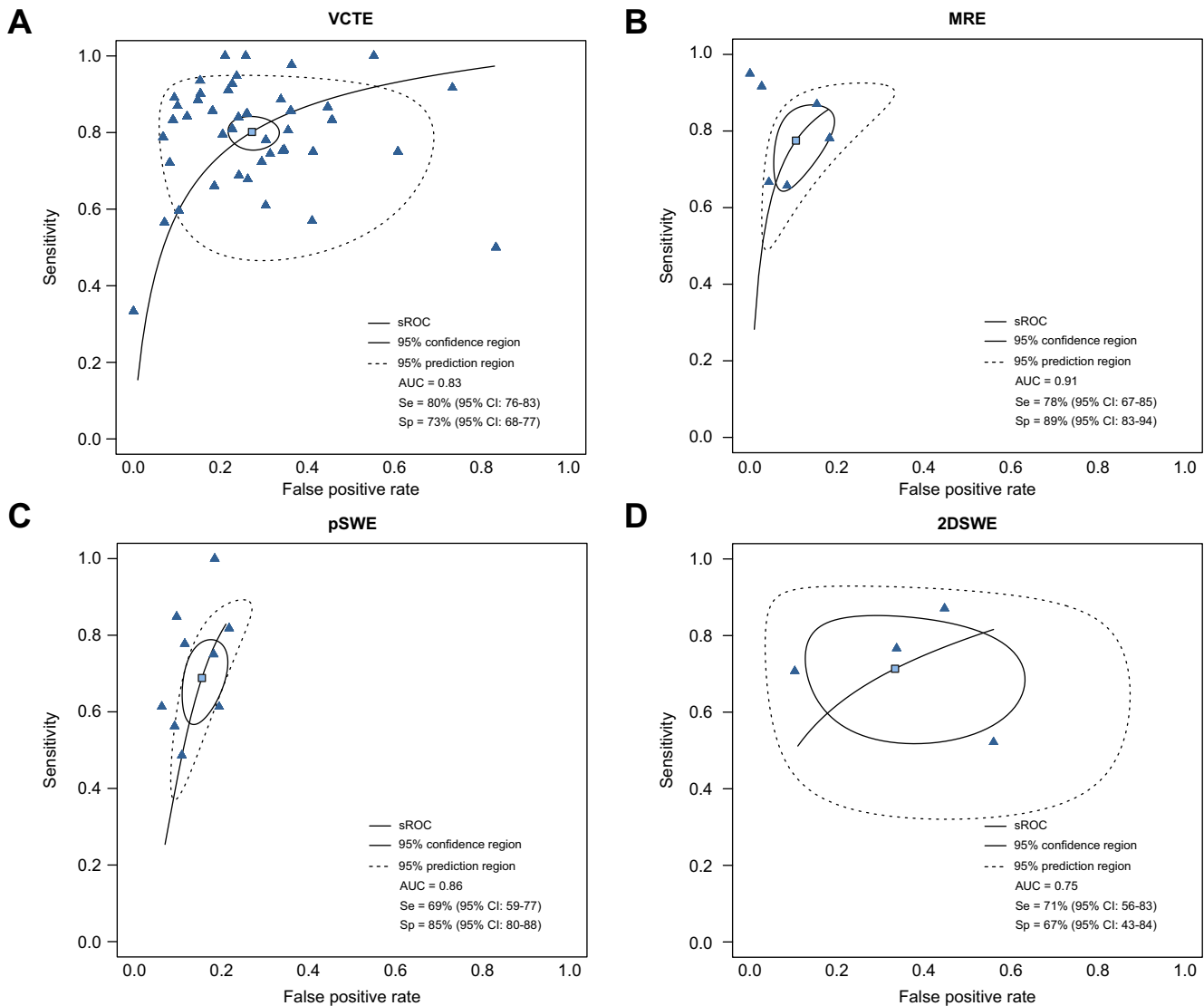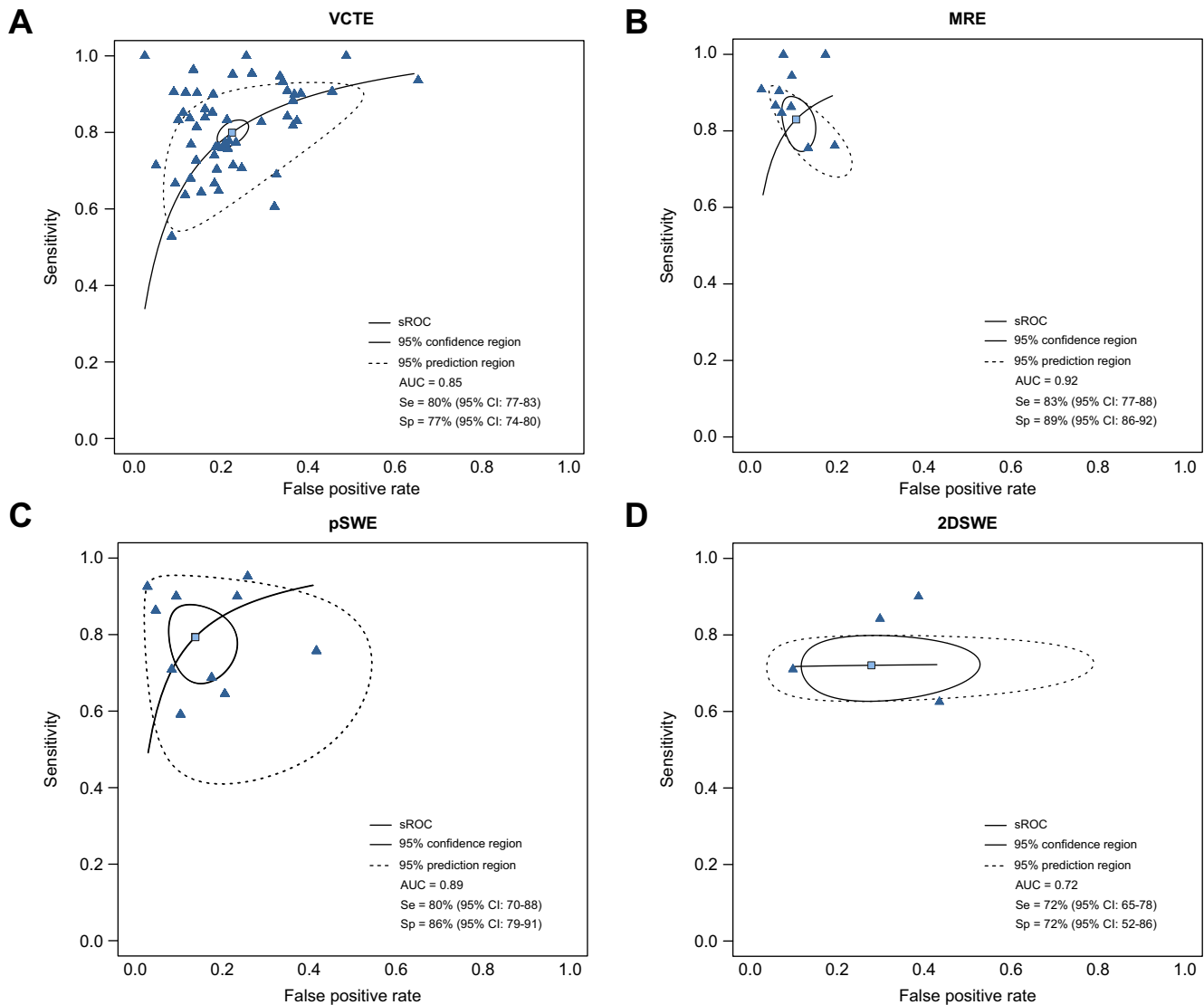
Fig. 3. sROC curves and test performance to detect significant fibrosis (F0-1 *vs.* F2-4). A bivariate logit-normal random effects model was used to estimate the mean sensitivity, and mean specificity for liver stiffness measured using (A) VCTE, (B) MRE, (C) pSWE and (D) 2DSWE. 2DSWE, 2-dimensional shear wave elastography; MRE, magnetic resonance elastography; pSWE, point shear wave elastography; sROC, summary receiver operating characteristic; VCTE, vibration-controlled transient elastography.

rates of index tests. These gaps in the literature are reflected in current society guidelines which suggest that elastography tests could be used as screening tests for advanced fibrosis, but with no specific cut-off recommendations for this context.[105]

Furthermore, the available data address the use of these index tests exclusively in the context of the diagnosis of the target conditions in secondary care. Data for their performance in other contexts (*e.g.* screening for a target condition in primary care, or using an index test to indicate prognosis or predict treatment response) are therefore needed. In addition, for the 2 other ultrasound elastography techniques (pSWE and 2DSWE) there is also a lack of data on validated reliability criteria.

Another area where data are lacking is the diagnostic performance of these tests for NASH. While traditionally the modalities we examined in this review were developed for fibrosis assessment, it is becoming increasingly important to be able to

identify patients with significant fibrosis and NASH, as these patients are thought to be more likely to benefit from pharmacological treatments. Several recent studies have sought to address this using combination index tests,[106–108] but as we only included single index tests here, we did not evaluate these.

Beyond diagnostic accuracy, the clinical utility of any non-invasive biomarker needs to also consider the cost-effectiveness for the healthcare provider, the opportunity costs for the patient (*e.g.* costs of scheduling an appointment and costs of travel to the appointment), and availability of alternative options. The equipment used for VCTE acquisitions can only be used for the purpose of liver stiffness measurement, but it is a point-of-care test. On the other hand, pSWE and 2DSWE may be performed by technicians trained in shear wave elastography and are incorporated into conventional ultrasound machines that can be used for alternative purposes beyond liver stiffness
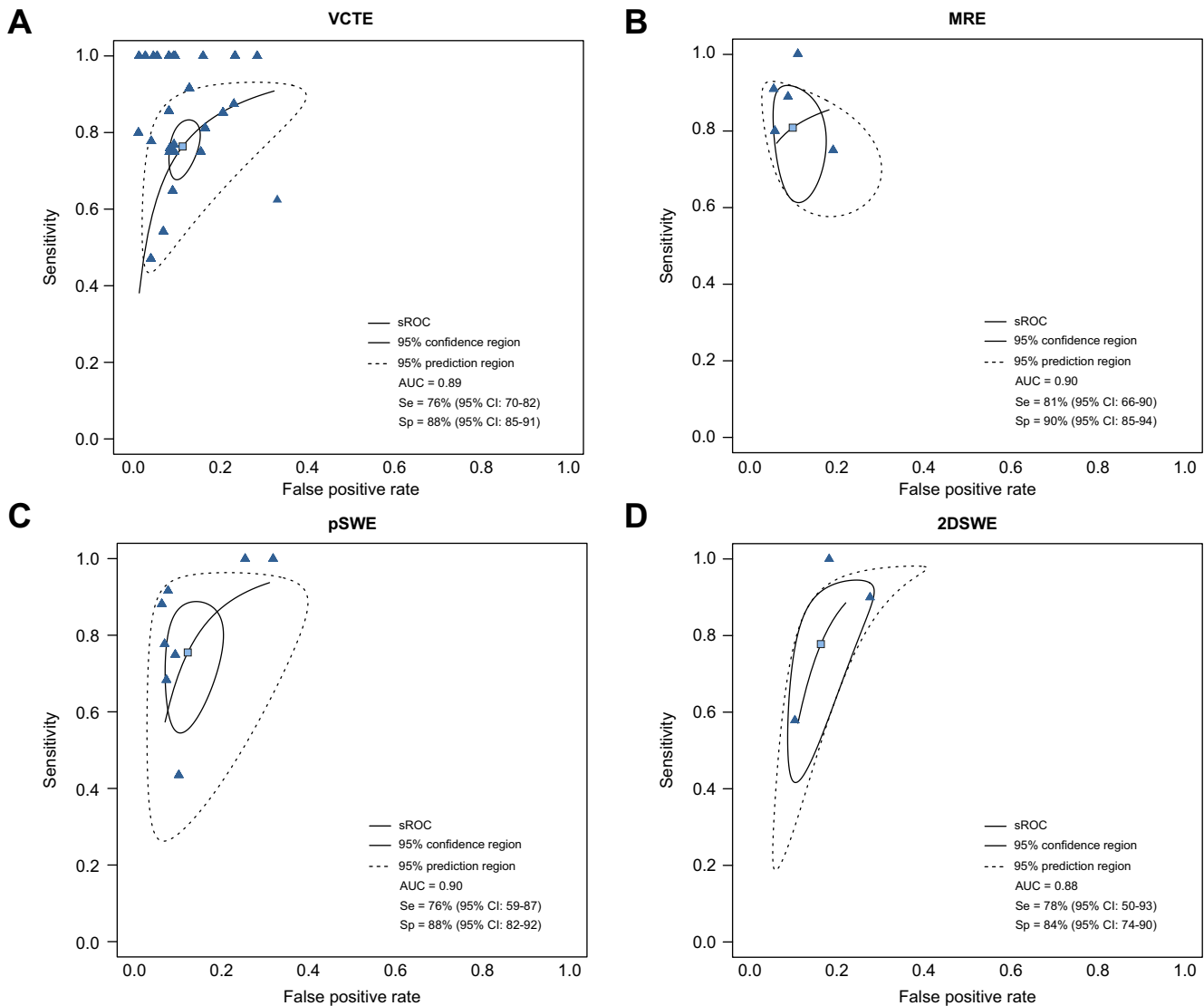
Fig. 4. sROC curves and test performance to detect advanced fibrosis (F0-2 *vs.* F3-4). A bivariate logit-normal random effects model was used to estimate the mean sensitivity, and mean specificity for liver stiffness measured using (A) VCTE, (B) MRE, (C) pSWE and (D) 2DSWE. 2DSWE, 2-dimensional shear wave elastography; MRE, magnetic resonance elastography; pSWE, point shear wave elastography; sROC, summary receiver operating characteristic; VCTE, vibration-controlled transient elastography.

measurement. MRE requires additional hardware and therefore has costs associated with installation, but reporting can be done by the local radiologists without any additional costs thereafter. On the other hand, LMS and deMILI can be performed in scanners with certain specifications and have fewer installation costs, but require a post-acquisition reporting service with a per-scan cost. Some data on the cost-effectiveness of LMS in combination with FibroScan® are available,[30,109] but a health economics analysis was beyond the aims of this systematic review and may need to be considered in the future.

We acknowledge several limitations of our study. We were able to evaluate diagnostic accuracy at different thresholds in only 6 studies reporting VCTE findings, as most studies reported only a single cut-off value. Having only a few studies in our multiple-threshold analysis also prevented us from proposing

cut-off values for pre-defined levels of sensitivity and specificity. However, we note that in the studies included in the multiple-threshold meta-analysis, PPV was low when considering prevalences typically seen in routine practice (prevalence <30%). Furthermore, statistical comparison between the imaging biomarkers was not possible because of the inadequate number of studies that examined these biomarkers contemporaneously in the same study population. Whilst some studies reported subgroup analysis for patients with increased BMI, the impact of other potential confounding factors such as abdominal wall thickness, inflammation and steatosis were not explored in our study. A time interval of up to 6 months between index test and liver biopsy allowed us to include a large number of studies in our meta-analysis. However, regression of steatosis and fibrosis can occur within 6 months, particularly if patients have had

Fig. 5. sROC curves and test performance to detect cirrhosis (F0-3 *vs.* F4). A bivariate logit-normal random effects model was used to estimate the mean sensitivity, and mean specificity for liver stiffness measured using (A) VCTE, (B) MRE, (C) pSWE and (D) 2DSWE. 2DSWE, 2-dimensional shear wave elastography; MRE, magnetic resonance elastography; pSWE, point shear wave elastography; sROC, summary receiver operating characteristic; VCTE, vibration-controlled transient elastography.
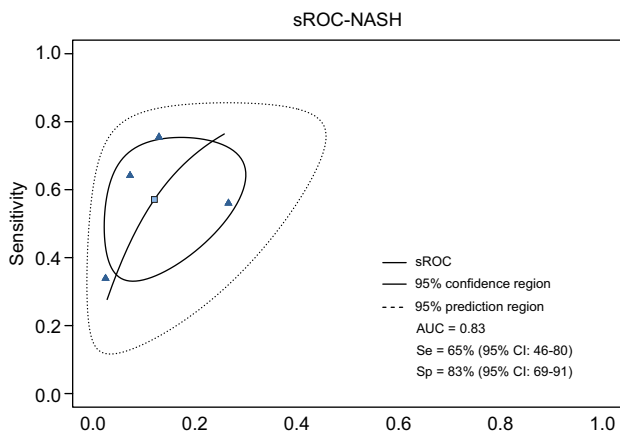
Fig. 6. sROC curves (bivariate logit-normal model) for the diagnostic accuracy of MRE to detect NASH (simple steatosis *vs.* NASH). MRE, magnetic resonance elastography; NASH, non-alcoholic steatohepatitis; sROC, summary receiver operating characteristic.

significant weight loss in the intervening period. Nevertheless, our sensitivity analysis showed that there was no difference in diagnostic accuracy when considering only studies that included paired biopsy and index biomarker within a 3-month interval. The patient population recruited to the included studies was not completely uniform and included patients with biopsy-proven NAFLD (from cohorts who underwent biopsy to evaluate known or suspected NAFLD), patients from exclusively bariatric cohorts and NAFLD sub-populations from cohorts of patients with mixed liver disease aetiologies. These factors need to be taken into consideration when interpreting the median prevalence and diagnostic accuracy data reported in our study.

In conclusion, in patients with NAFLD where liver stiffness can be measured successfully, VCTE, MRE, pSWE and 2DSWE have a good diagnostic accuracy for the assessment of fibrosis, but only MRE and pSWE meet the minimum acceptable criteria of at least 80% sensitivity and specificity for the diagnosis of advanced fibrosis. These promising results however, are likely to

be overestimates of the true diagnostic accuracy as intention-to-diagnose analyses and validation of pre-specified cut-offs are lacking from the literature. Future studies, like the LITMUS Imaging Study being conducted in Europe and the USA currently, should also evaluate the newer 2DSWE and MRI techniques, and provide data on head-to-head comparisons of the various techniques.

Abbreviations

2DSWE, two-dimensional shear wave elastography; cT1, iron-corrected T1 relaxation time; deMILI, detection of metabolic liver injury; DWI, diffusion-weighted imaging; LMS, Liver-*MultiScan*™; MRE, magnetic resonance elastography; MRI, magnetic resonance imaging; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; NPV, negative predictive value; PPV, positive predictive value; pSWE, point shear wave elastography; sAUC, summary area under the curve; sROC, summary receiver operating characteristic curve; VCTE, vibration-controlled transient elastography.

Data availability statement

This study has not generated any primary data. The data for this study were extracted from previously published manuscripts. Therefore, the data are already freely available in the manuscripts of the published studies which were included in this systematic review.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhep.2021.04.044.

References

[1] 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;64:73–84.

[2] Anstee QM, Targher G, Day CP. Progression of NAFLD to diabetes mellitus, cardiovascular disease or cirrhosis. Nat Rev Gastroenterol Hepatol 2013;10:330–344.

[3] 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;62:1148–1155.

[4] Angulo P, Kleiner DE, Dam-Larsen S, Adams LA, Bjornsson ES, Charatcharoenwitthaya P, et al. Liver fibrosis, but No other histologic features, is associated with long-term outcomes of patients with

nonalcoholic fatty liver disease. Gastroenterology 2015;149:389–397 e310.

[5] 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;65:1557–1565.

[6] Hagstrom H, Nasr P, Ekstedt M, Hammar U, Stal P, Hultcrantz R, et al. Fibrosis stage but not NASH predicts mortality and time to development of severe liver disease in biopsy-proven NAFLD. J Hepatol 2017;67:1265–1273.

[7] Castera L, Friedrich-Rust M, Loomba R. Noninvasive assessment of liver disease in patients with nonalcoholic fatty liver disease. Gastroenterology 2019;156. 1264-1281.e1264.

[8] 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;41:1313–1321.

[9] 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;155:529–536.

[10] Vali Y, Lee J, Boursier J, Spijker R, Löffler J, Verheij J, et al. Enhanced liver fibrosis test for the non-invasive diagnosis of fibrosis in patients with NAFLD: a systematic review and meta-analysis. J Hepatol 2020;73:252–262.

[11] Steinhauser S, Schumacher M, Rücker G. Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. BMC Med Res Methodol 2016;16.

[12] Schneider A, Linde K, Reitsma JB, Steinhauser S, Rücker G. A novel statistical model for analyzing data of a systematic review generates optimal cutoff values for fractional exhaled nitric oxide for asthma diagnosis. J Clin Epidemiol 2017;92:69–78.

[13] Burkner PC, Doebler P. Testing for publication bias in diagnostic meta-analysis: a simulation study. Stat Med 2014;33:3061–3077.

[14] Doebler P. mada: meta-analysis of diagnostic accuracy. 2019 [cited 2020 25/08/2020]; R package version 0.5.9]. Available from: https://CRAN.R-project.org/package=mada.

[15] Rücker G, Steinhauser S. diagmeta: meta-analysis of diagnostic accuracy studies with several cutpoints. 2020.

[16] Agrawal S, Hoad CL, Francis ST, Guha IN, Kaye P, Aithal GP. Visual morphometry and three non-invasive markers in the evaluation of liver fibrosis in chronic liver disease. Scand J Gastroenterol 2017;52:107–115.

[17] Al Juboori A, Nasser M, Hammoud GM, Rector RS, Parks EJ, Ibdah JA, et al. The utility of liver elastography in evaluation of liver fibrosis in obese patients undergoing bariatric surgery. Hepatology 2018;68.

[18] Anstee QM, Lawitz EJ, Alkhouri N, Wong VW, 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.

[19] Attia D, Bantel H, Lenzen H, Manns MP, Gebel MJ, Potthoff A. Liver stiffness measurement using acoustic radiation force impulse elastography in overweight and obese patients. Aliment Pharmacol Ther 2016;44:366–379.

[20] Aykut UE, Akyuz U, Yesil A, Eren F, Gerin F, Ergelen R, et al. A comparison of FibroMeter NAFLD Score, NAFLD fibrosis score, and transient elastography as noninvasive diagnostic tools for hepatic fibrosis in patients with biopsy-proven non-alcoholic fatty liver disease. Scand J Gastroenterol 2014;49:1343–1348.

[21] Boursier J, Vergniol J, Guillet A, Hiriart JB, Lannes A, Le Bail B, 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 2016;65:570–578.

[22] Boursier J, Guillaume M, Leroy V, Irles 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;71:389–396.

[23] Cardoso AC, Cravo C, Calcado FL, Rezende G, Campos CFF, Neto JMA, et al. The performance of M and XL probes of FibroScan for the diagnosis of steatosis and fibrosis on a Brazilian nonalcoholic fatty liver disease cohort. Eur J Gastroenterol Hepatol 2020;32:231–238.

[24] Cassinotto C, Lapuyade B, Ait-Ali A, Vergniol J, Gaye D, Foucher J, et al. Liver fibrosis: noninvasive assessment with acoustic radiation force impulse elastography–comparison with FibroScan M and XL probes and FibroTest in patients with chronic liver disease. Radiology 2013;269:283–292.

[25] Cassinotto C, Boursier J, de Ledinghen V, Lebigot J, Lapuyade B, Cales P, et al. Liver stiffness in nonalcoholic fatty liver disease: a comparison of supersonic shear imaging, FibroScan, and ARFI with liver biopsy. Hepatology 2016;63:1817–1827.

[26] Chan WK, Nik Mustapha NR, Mahadeva S. A novel 2-step approach combining the NAFLD fibrosis score and liver stiffness measurement for predicting advanced fibrosis. Hepatol Int 2015;9:594–602.

[27] Chan WK, Nik Mustapha NR, Wong GL, Wong VW, Mahadeva S. Controlled attenuation parameter using the FibroScan(R) XL probe for quantification of hepatic steatosis for non-alcoholic fatty liver disease in an Asian population. United Eur Gastroenterol J 2017;5:76–85.

[28] Clet M, Miette V, Eddowes P, Allison MED, Tsochatzis EA, Anstee QM, et al. Expanding the Use of the VCTE XL probe in morbid obese patients: validation of a new automated adaptive measurement depths algorithm in a large UK multicenter cohort. Hepatology 2018;68.

[29] Das K, Sarkar R, Ahmed SM, Mridha AR, Mukherjee PS, Das K, et al. "Normal" liver stiffness measure (LSM) values are higher in both lean and obese individuals: a population-based study from a developing country. Hepatology 2012;55:584–593.

[30] 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 2018;47:631–644.

[31] Eddowes PJ, Sasso M, Allison M, Tsochatzis E, Anstee QM, Sheridan D, et al. Accuracy of FibroScan controlled attenuation parameter and liver stiffness measurement in assessing steatosis and fibrosis in patients with nonalcoholic fatty liver disease. Gastroenterology 2019;156:1717–1730.

[32] Ergelen R, Yilmaz Y, Asedov R, Celikel C, Akin H, Bugdayci O, et al. Comparison of Doppler ultrasound and transient elastography in the diagnosis of significant fibrosis in patients with nonalcoholic steatohepatitis. Abdom Radiol (NY) 2016;41:1505–1510.

[33] Forlano R, Manousou P, Mullish B, Olaoke A, Khelifa MZ, Taylor-Robinson SD, et al. Assessment of non invasive markers of fibrosis against collagen quantitation and NASH-CRN scoring in liver biopsies of NAFLD patients. Hepatology 2017;66.

[34] Gaia S, Carenzi S, Barilli AL, Bugianesi E, Smedile A, Brunello F, et al. Reliability of transient elastography for the detection of fibrosis in non-alcoholic fatty liver disease and chronic viral hepatitis. J Hepatol 2011;54:64–71.

[35] Gaia S, Campion D, Evangelista A, Spandre M, Cosso L, Brunello F, et al. Non-invasive score system for fibrosis in chronic hepatitis: proposal for a model based on biochemical, FibroScan and ultrasound data. Liver Int 2015;35:2027–2035.

[36] Gallego-Duran R, Cerro-Salido P, Gomez-Gonzalez E, Pareja MJ, Ampuero J, Rico MC, et al. Imaging biomarkers for steatohepatitis and fibrosis detection in non-alcoholic fatty liver disease. Sci Rep 2016;6:31421.

[37] Garg H, Aggarwal S, Shalimar, Yadav R, Datta Gupta S, Agarwal L, et al. Utility of transient elastography (fibroscan) and impact of bariatric surgery on nonalcoholic fatty liver disease (NAFLD) in morbidly obese patients. Surg Obes Relat Dis 2018;14:81–91.

[38] Hee E, Ching H, Jeffrey GP, Adams L, Kemp WW, Roberts SK, et al. Comparison of NAFLD Fibrosis Score (NFS) and transient elastography (Fibroscan), alone and in combination, as non-invasive methods for the evaluation of fibrosis in non-alcoholic fatty liver disease (NAFLD). J Gastroenterol Hepatol 2014;29:76.

[39] Imajo K, Kessoku T, Honda Y, Tomeno W, Ogawa Y, Mawatari H, et al. Magnetic resonance imaging more accurately classifies steatosis and fibrosis in patients with nonalcoholic fatty liver disease than transient elastography. Gastroenterology 2016;150. 626-637.e627.

[40] Inadomi C, Takahashi H, Ogawa Y, Oeda S, Imajo K, Kubotsu Y, et al. Accuracy of the Enhanced Liver Fibrosis test, and combination of the Enhanced Liver Fibrosis and non-invasive tests for the diagnosis of advanced liver fibrosis in patients with non-alcoholic fatty liver disease. Hepatol Res 2020;50:682–692.

[41] Kao WY, Chang IW, Chen CL, Su CW, Fang SU, Tang JH, et al. Fibroscan-based score to predict significant liver fibrosis in morbidly obese patients with nonalcoholic fatty liver disease. Obes Surg 2020;30:1249–1257.

[42] 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: e0141649.

[43] Kumar R, Rastogi A, Sharma MK, Bhatia V, Tyagi P, Sharma P, et al. Liver stiffness measurements in patients with different stages of nonalcoholic fatty liver disease: diagnostic performance and clinicopathological correlation. Dig Dis Sci 2013;58:265–274.

[44] Kwok R, Choi KC, Wong GL, Zhang Y, Chan HL, Luk AO, et al. Screening diabetic patients for non-alcoholic fatty liver disease with controlled attenuation parameter and liver stiffness measurements: a prospective cohort study. Gut 2016;65:1359–1368.

[45] Labenz C, Huber Y, Kalliga E, Nagel M, Ruckes C, Straub BK, et al. Predictors of advanced fibrosis in non-cirrhotic non-alcoholic fatty liver disease in Germany. Aliment Pharmacol Ther 2018;48:1109–1116.

[46] Lee HW, Park SY, Kim SU, Jang JY, Park H, Kim JK, et al. Discrimination of nonalcoholic steatohepatitis using transient elastography in patients with nonalcoholic fatty liver disease. PloS One 2016;11:e0157358.

[47] Lee MS, Bae JM, Joo SK, Woo H, Lee DH, Jung YJ, et al. Prospective comparison among transient elastography, supersonic shear imaging, and ARFI imaging for predicting fibrosis in nonalcoholic fatty liver disease. PloS One 2017;12:e0188321.

[48] Lee JI, Lee HW, Lee KS. Value of controlled attenuation parameter in fibrosis prediction in nonalcoholic steatohepatitis. World J Gastroenterol 2019;25:4959–4969.

[49] Lee YS, Yoo YJ, Jung YK, Kim JH, Seo YS, Yim HJ, et al. Multiparametric MR is a valuable modality for evaluating disease severity of nonalcoholic fatty liver disease. Clin Transl Gastroenterol 2020;11:e00157.

[50] Leong WL, Lai LL, Nik Mustapha NR, Vijayananthan A, Rahmat K, Mahadeva S, et al. Comparing point shear wave elastography (ElastPQ) and transient elastography for diagnosis of fibrosis stage in non-alcoholic fatty liver disease. J Gastroenterol Hepatol 2020;35:135–141.

[51] Loong TC, Wei JL, Leung JC, Wong GL, Shu SS, Chim AM, et al. Application of the combined FibroMeter vibration-controlled transient elastography algorithm in Chinese patients with non-alcoholic fatty liver disease. J Gastroenterol Hepatol 2017;32:1363–1369.

[52] Lupsor M, Badea R, Stefanescu H, Grigorescu M, Serban A, Radu C, et al. Performance of unidimensional transient elastography in staging non-alcoholic steatohepatitis. J Gastrointestin Liver Dis 2010;19:53–60.

[53] Mahadeva S, Mahfudz AS, Vijayanathan A, Goh KL, Kulenthran A, Cheah PL. Performance of transient elastography (TE) and factors associated with discordance in non-alcoholic fatty liver disease. J Dig Dis 2013;14:604–610.

[54] Morrison M, Spataro J, Claridge L, Puri P, Hughes H, Shetty A, et al. Accuracy of fibroscan continuous attenuation parameter and liver stiffness in assessing steatosis and fibrosis in USA veterans with nonalcoholic fatty liver disease. Gastroenterology 2020;158. S237-S237.

[55] Myers RP, Pomier-Layrargues G, Kirsch R, Pollett A, Duarte-Rojo A, Wong D, et al. Feasibility and diagnostic performance of the FibroScan XL probe for liver stiffness measurement in overweight and obese patients. Hepatology 2012;55:199–208.

[56] Naveau S, Lamouri K, Pourcher G, Njike-Nakseu M, Ferretti S, Courie R, et al. The diagnostic accuracy of transient elastography for the diagnosis of liver fibrosis in bariatric surgery candidates with suspected NAFLD. Obes Surg 2014;24:1693–1701.

[57] Naveau S, Voican CS, Lebrun A, Gaillard M, Lamouri K, Njike-Nakseu M, et al. Controlled attenuation parameter for diagnosing steatosis in bariatric surgery candidates with suspected nonalcoholic fatty liver disease. Eur J Gastroenterol Hepatol 2017;29:1022–1030.

[58] Oeda S, Takahashi H, Imajo K, Seko Y, Ogawa Y, Moriguchi M, et al. Accuracy of liver stiffness measurement and controlled attenuation parameter using FibroScan® M/XL probes to diagnose liver fibrosis and steatosis in patients with nonalcoholic fatty liver disease: a multicenter prospective study. J Gastroenterol 2020;55:428–440.

[59] Okajima A, Sumida Y, Taketani H, Hara T, Seko Y, Ishiba H, et al. Liver stiffness measurement to platelet ratio index predicts the stage of liver fibrosis in non-alcoholic fatty liver disease. Hepatol Res 2017;47:721–730.

[60] Ooi GJ, Earnest A, Kemp WW, Burton PR, Laurie C, Majeed A, et al. Evaluating feasibility and accuracy of non-invasive tests for nonalcoholic fatty liver disease in severe and morbid obesity. Int J Obes (Lond) 2018;42:1900–1911.

[61] Pais R, Lebray P, Fedchuk L, Charlotte F, Poynard T, Ratziu V. Validation of a simple algorithm combining serum markers and elastometry for the diagnosis of fibrosis in NAFLD. Hepatology 2011;54.

[62] Park CC, Nguyen P, Hernandez C, Bettencourt R, Ramirez K, Fortney L, et al. Magnetic resonance elastography vs. Transient elastography in detection of fibrosis and noninvasive measurement of steatosis in patients with biopsy-proven nonalcoholic fatty liver disease. Gastroenterology 2017;152. 598-607.e592.

[63] Pavlides M, Banerjee R, Tunnicliffe EM, Kelly C, Collier J, Wang LM, et al. Multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease severity. Liver Int 2017;37:1065–1073.

[64] Petta S, Di Marco V, Camma C, Butera G, Cabibi D, Craxi A. Reliability of liver stiffness measurement in non-alcoholic fatty liver disease: the effects of body mass index. Aliment Pharmacol Ther 2011;33:1350–1360.

[65] Petta S, Vanni E, Bugianesi E, Di Marco V, Camma C, Cabibi D, et al. The combination of liver stiffness measurement and NAFLD fibrosis score improves the noninvasive diagnostic accuracy for severe liver fibrosis in patients with nonalcoholic fatty liver disease. Liver Int 2015;35:1566–1573.

[66] Petta S, Maida M, Macaluso FS, Di Marco V, Camma C, Cabibi D, et al. The severity of steatosis influences liver stiffness measurement in patients with nonalcoholic fatty liver disease. Hepatology 2015;62:1101–1110.

[67] Petta S, Wong VW, Camma C, Hiriart JB, Wong GL, Vergniol J, et al. Serial combination of non-invasive tools improves the diagnostic accuracy of severe liver fibrosis in patients with NAFLD. Aliment Pharmacol Ther 2017;46:617–627.

[68] Petta S, Wong VW, Camma C, Hiriart JB, Wong GL, Marra F, et al. Improved noninvasive prediction of liver fibrosis by liver stiffness measurement in patients with nonalcoholic fatty liver disease accounting for controlled attenuation parameter values. Hepatology 2017;65:1145–1155.

[69] Rosso C, Caviglia GP, Abate ML, Vanni E, Mezzabotta L, Touscoz GA, et al. Cytokeratin 18-Aspartate396 apoptotic fragment for fibrosis detection in patients with non-alcoholic fatty liver disease and chronic viral hepatitis. Dig Liver Dis 2016;48:55–61.

[70] Seki K, Shima T, Oya H, Mitsumoto Y, Mizuno M, Okanoue T. Assessment of transient elastography in Japanese patients with non-alcoholic fatty liver disease. Hepatol Res 2017;47:882–889.

[71] Shen F, Zheng RD, Shi JP, Mi YQ, Chen GF, Hu X, et al. Impact of skin capsular distance on the performance of controlled attenuation parameter in patients with chronic liver disease. Liver Int 2015;35:2392–2400.

[72] Shima T, Sakai K, Oya H, Katayama T, Mitsumoto Y, Mizuno M, et al. Diagnostic accuracy of combined biomarker measurements and vibration-controlled transient elastography (VCTE) for predicting fibrosis stage of non-alcoholic fatty liver disease. J Gastroenterol 2020;55:100–112.

[73] 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;7:1113–1123.

[74] Tapper EB, Challies T, Nasser I, Afdhal NH, Lai M. The performance of vibration controlled transient elastography in a US cohort of patients with nonalcoholic fatty liver disease. Am J Gastroenterol 2016;111:677–684.

[75] Wong VW, Vergniol J, Wong GL, Foucher J, Chan HL, Le Bail B, et al. Diagnosis of fibrosis and cirrhosis using liver stiffness measurement in nonalcoholic fatty liver disease. Hepatology 2010;51:454–462.

[76] Wong VW, Vergniol J, Wong GL, Foucher J, Chan AW, Chermak F, et al. Liver stiffness measurement using XL probe in patients with nonalcoholic fatty liver disease. Am J Gastroenterol 2012;107:1862–1871.

[77] Wong VW, Irles M, Wong GL, Shili S, Chan AW, Merrouche W, et al. Unified interpretation of liver stiffness measurement by M and XL probes in non-alcoholic fatty liver disease. Gut 2019;68:2057–2064.

[78] Yoneda M, Yoneda M, Mawatari H, Fujita K, Endo H, Iida H, et al. Noninvasive assessment of liver fibrosis by measurement of stiffness in patients with nonalcoholic fatty liver disease (NAFLD). Dig Liver Dis 2008;40:371–378.

[79] Younes R, Rosso C, Petta S, Cucco M, Marietti M, Caviglia GP, et al. Usefulness of the index of NASH - ION for the diagnosis of steatohepatitis in patients with non-alcoholic fatty liver: an external validation study. Liver Int 2018;38:715–723.

[80] Ziol M, Kettaneh A, Ganne-Carrie N, Barget N, Tengher-Barna I, Beaugrand M. Relationships between fibrosis amounts assessed by morphometry and liver stiffness measurements in chronic hepatitis or steatohepatitis. Eur J Gastroenterol Hepatol 2009;21:1261–1268.

[81] Chen J, Talwalkar JA, Yin M, Glaser KJ, Sanderson SO, Ehman RL. Early detection of nonalcoholic steatohepatitis in patients with nonalcoholic fatty liver disease by using MR elastography. Radiology 2011;259:749–756.

[82] Costa-Silva L, Ferolla SM, Lima AS, Vidigal PVT, Ferrari TCA. MR elastography is effective for the non-invasive evaluation of fibrosis and necroinflammatory activity in patients with nonalcoholic fatty liver disease. Eur J Radiol 2018;98:82–89.

[83] Cui J, Ang B, Haufe W, Hernandez C, Verna EC, Sirlin CB, et al. Comparative diagnostic accuracy of magnetic resonance elastography vs.. eight clinical prediction rules for non-invasive diagnosis of advanced fibrosis in biopsy-proven non-alcoholic fatty liver disease: a prospective study. Aliment Pharmacol Ther 2015;41:1271–1280.

[84] Cui J, Heba E, Hernandez C, Haufe W, Hooker J, Andre MP, et al. Magnetic resonance elastography is superior to acoustic radiation force impulse for the Diagnosis of fibrosis in patients with biopsy-proven nonalcoholic fatty liver disease: a prospective study. Hepatology 2016;63:453–461.

[85] Kim D, Kim WR, Talwalkar JA, Kim HJ, Ehman RL. Advanced fibrosis in nonalcoholic fatty liver disease: noninvasive assessment with MR elastography. Radiology 2013;268:411–419.

[86] Kim JW, Lee YS, Park YS, Kim BH, Lee SY, Yeon JE, et al. Multiparametric MR index for the diagnosis of non-alcoholic steatohepatitis in patients with non-alcoholic fatty liver disease. Sci Rep 2020;10:2671.

[87] Loomba R, Lam J, Wolfson T, Ang B, Bhatt A, Peterson MR, et al. Magnetic resonance elastography accurately predicts advanced fibrosis in NAFLD: a pilot study with paired-liver biopsy. Gastroenterology 2013;144. S1012-S1012.

[88] Loomba R, Wolfson T, Ang B, Hooker J, Behling C, Peterson M, et al. Magnetic resonance elastography predicts advanced fibrosis in patients with nonalcoholic fatty liver disease: a prospective study. Hepatology 2014;60:1920–1928.

[89] Loomba R, Cui J, Wolfson T, Haufe W, Hooker J, Szeverenyi N, et al. Novel 3D magnetic resonance elastography for the noninvasive diagnosis of advanced fibrosis in NAFLD: a prospective study. Am J Gastroenterol 2016;111:986–994.

[90] Fierbinteanu Braticevici C, Sporea I, Panaitescu E, Tribus L. Value of acoustic radiation force impulse imaging elastography for non-invasive evaluation of patients with nonalcoholic fatty liver disease. Ultrasound Med Biol 2013;39:1942–1950.

[91] Joo SK, Kim W, Kim D, Kim JH, Oh S, Lee KL, et al. Steatosis severity affects the diagnostic performances of noninvasive fibrosis tests in nonalcoholic fatty liver disease. Liver Int 2018;38:331–341.

[92] Medellin A, Pridham G, Urbanski SJ, Jayakumar S, Wilson SR. Acoustic radiation force impulse and conventional ultrasound in the prediction of cirrhosis complicating fatty liver: does body mass index independently alter the results? Ultrasound Med Biol 2019;45:3160–3171.

[93] Palmeri ML, Wang MH, Rouze NC, Abdelmalek MF, Guy CD, Moser B, et al. Noninvasive evaluation of hepatic fibrosis using acoustic radiation force-based shear stiffness in patients with nonalcoholic fatty liver disease. J Hepatol 2011;55:666–672.

[94] Zhang Da-kun CMYL. The diagnostic value of ARFI imaging and APRI for quantitative evaluating the degree of liver fibrosis in NAFLD patients. J Chin Clin Med Imaging 2014:186–187.

[95] Ozturk A, Mohammadi R, Pierce TT, Kamarthi S, Dhyani M, Grajo JR, et al. Diagnostic accuracy of shear wave elastography as a non-invasive biomarker of high-risk non-alcoholic steatohepatitis in patients with non-alcoholic fatty liver disease. Ultrasound Med Biol 2020;46:972–980.

[96] Takeuchi H, Sugimoto K, Oshiro H, Iwatsuka K, Kono S, Yoshimasu Y, et al. Liver fibrosis: noninvasive assessment using supersonic shear imaging and FIB4 index in patients with non-alcoholic fatty liver disease. J Med Ultrason (2001) 2018;45:243–249.

[97] Parente DB, Paiva FF, Oliveira Neto JA, Machado-Silva L, Figueiredo FA, Lanzoni V, et al. Intravoxel incoherent motion diffusion weighted MR imaging at 3.0 T: assessment of steatohepatitis and fibrosis compared with liver biopsy in type 2 diabetic patients. PloS One 2015;10:e0125653.

[98] Audière S, Miette V, Fournier C, Whitehead J, Paredes A, Sandrin L, et al. Continuous CAP algorithm: reduced variability in a prospective cohort. J Hepatol 2020;73:S436.

[99] Newsome PN, Clet M, Czernichow S, Eddowes P, Allison M, Tsochatzis E, et al. Expanding the use of the vibration controlled transient elastography in morbid obese patients: validation of a new automated adaptive measurement depths algorithm in a large pooled NAFLD cohort. J Hepatol 2020;73:S430–S431.

[100] Liang Y, Li D. Magnetic resonance elastography in staging liver fibrosis in non-alcoholic fatty liver disease: a pooled analysis of the diagnostic accuracy. BMC Gastroenterol 2020;20:89.

[101] Singh S, Venkatesh SK, Wang Z, Miller FH, Motosugi U, Low RN, et al. Diagnostic performance of magnetic resonance elastography in staging liver fibrosis: a systematic review and meta-analysis of individual participant data. Clin Gastroenterol Hepatol 2015;13. 440-451 e446.

[102] Xiao G, Zhu S, Xiao X, Yan L, Yang J, Wu G. Comparison of laboratory tests, ultrasound, or magnetic resonance elastography to detect fibrosis in patients with nonalcoholic fatty liver disease: a meta-analysis. Hepatology 2017;66:1486–1501.

[103] Besutti G, Valenti L, Ligabue G, Bassi MC, Pattacini P, Guaraldi G, et al. Accuracy of imaging methods for steatohepatitis diagnosis in non-alcoholic fatty liver disease patients: a systematic review. Liver Int 2019;39:1521–1534.

[104] Jiang W, Huang S, Teng H, Wang P, Wu M, Zhou X, et al. Diagnostic accuracy of point shear wave elastography and transient elastography for staging hepatic fibrosis in patients with non-alcoholic fatty liver disease: a meta-analysis. BMJ Open 2018;8:e021787.

[105] EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. J Hepatol 2016;64:1388–1402.

[106] Newsome PN, Sasso M, Deeks JJ, Paredes A, Boursier J, Chan WK, et al. FibroScan-AST (FAST) score for the non-invasive identification of patients with non-alcoholic steatohepatitis with significant activity and fibrosis: a prospective derivation and global validation study. Lancet Gastroenterol Hepatol 2020;5:362–373.

[107] Harrison SA, Ratziu V, Boursier J, Francque S, Bedossa P, Majd Z, et al. A blood-based biomarker panel (NIS4) for non-invasive diagnosis of non-alcoholic steatohepatitis and liver fibrosis: a prospective derivation and global validation study. Lancet Gastroenterol Hepatol 2020.

[108] Dennis A, Mouchti S, Kelly M, Fallowfield JA, Hirschfield G, Pavlides M, et al. A composite biomarker using multiparametric magnetic resonance imaging and blood analytes accurately identifies patients with non-alcoholic steatohepatitis and significant fibrosis. Sci Rep 2020;10:15308.

[109] Blake L, Duarte RV, Cummins C. Decision analytic model of the diagnostic pathways for patients with suspected non-alcoholic fatty liver disease using non-invasive transient elastography and multiparametric magnetic resonance imaging. BMJ open 2016;6:e010507.