Received: 3 March 2023 | Revised: 7 February 2024 | Accepted: 17 March 2024

DOI: 10.1111/liv.15914

Liver INTERNATIONAL WILEY

ORIGINAL ARTICLE

Diagnostic accuracy of non-invasive tests to screen for at-risk MASH—An individual participant data meta-analysis

Ferenc E. Mózes[1] | Jenny A. Lee[2] | Yasaman Vali[2] | Emmanuel A. Selvaraj[1,3,4] | Arjun N. A. Jayaswal[1] | Jérôme Boursier[5,6] | Victor de Lédinghen[7,8] | Monica Lupșor-Platon[9] | Yusuf Yilmaz[10,11] | Wah-Kheong Chan[12] | Sanjiv Mahadeva[12] | Thomas Karlas[13] | Johannes Wiegand[13] | Shalimar[14] | Emmanouil Tsochatzis[15] | Antonio Liguori[15,16] | Vincent Wai-Sun Wong[17] | Dae Ho Lee[18] | Adriaan G. Holleboom[19] | Anne-Marieke van Dijk[19] | Anne Linde Mak[19] | Hannes Hagström[20,21] | Camilla Akbari[21] | Masashi Hirooka[22] | Dong Hyeon Lee[23] | Won Kim[23] | Takeshi Okanoue[24] | Toshihide Shima[24] | Atsushi Nakajima[25] | Masato Yoneda[25] | Paul J. Thuluvath[26,27] | Feng Li[26] | Annalisa Berzigotti[28,29] | Yuly P. Mendoza[28,30] | Mazen Noureddin[31] | Emily Truong[32] | Céline Fournier-Poizat[33] | Andreas Geier[34] | Theresa Tuthill[35] | Carla Yunis[36] | Quentin M. Anstee[37,38] | Stephen A. Harrison[1] | Patrick M. Bossuyt[2] | Michael Pavlides[1,3,4]

Correspondence
Michael Pavlides, Oxford Centre for Clinical Magnetic Resonance Research (OCMR), Level 0, John Radcliffe Hospital, Headley Way, Oxford OX3 9DU, UK.
Email: michael.pavlides@cardiov.ox.ac.uk

Funding information
Innovative Medicines Initiative 2, Grant/Award Number: 777 377; European Union's Horizon 2020; Wellcome Trust; Royal Society, Grant/Award Number: 221805/Z/20/Z

Handling Editor: Dr. Luca Valenti

Abstract

Background & Aims: There is a need to reduce the screen failure rate (SFR) in metabolic dysfunction-associated steatohepatitis (MASH) clinical trials (MASH+F2-3; MASH+F4) and identify people with high-risk MASH (MASH+F2-4) in clinical practice. We aimed to evaluate non-invasive tests (NITs) screening approaches for these target conditions.

Methods: This was an individual participant data meta-analysis for the performance of NITs against liver biopsy for MASH+F2-4, MASH+F2-3 and MASH+F4. Index tests were the FibroScan-AST (FAST) score, liver stiffness measured using vibration-controlled transient elastography (LSM-VCTE), the fibrosis-4 score (FIB-4) and the NAFLD fibrosis score (NFS). Area under the receiver operating characteristics curve (AUROC) and thresholds including those that achieved 34% SFR were reported.

Results: We included 2281 unique cases. The prevalence of MASH+F2-4, MASH+F2-3 and MASH+F4 was 31%, 24% and 7%, respectively. Area under the receiver operating characteristics curves for MASH+F2-4 were .78, .75, .68 and .57 for FAST, LSM-VCTE, FIB-4 and NFS. Area under the receiver operating characteristics curves for

For affiliations refer to page 12.

MASH+F2-3 were .73, .67, .60, .58 for FAST, LSM-VCTE, FIB-4 and NFS. Area under the receiver operating characteristics curves for MASH+F4 were .79, .84, .81, .76 for FAST, LSM-VCTE, FIB-4 and NFS. The sequential combination of FIB-4 and LSM-VCTE for the detection of MASH+F2-3 with threshold of .7 and 3.48, and 5.9 and 20 kPa achieved SFR of 67% and sensitivity of 60%, detecting 15 true positive cases from a theoretical group of 100 participants at the prevalence of 24%.

Conclusions: Sequential combinations of NITs do not compromise diagnostic performance and may reduce resource utilisation through the need of fewer LSM-VCTE examinations.

KEYWORDS
at-risk MASH, FAST, FIB-4, LSM-VCTE, MASH, NFS, non-invasive tests

1 | INTRODUCTION

Metabolic dysfunction-associated steatotic liver disease (MASLD), previously known as non-alcoholic fatty liver disease (NAFLD)[1] is the most common cause of chronic liver disease worldwide, affecting 25%–30% of people in the general population.[2] In those with obesity and type 2 diabetes mellitus (T2DM), the prevalence of MASLD can be up to 70%.[2] Metabolic dysfunction-associated steatotic liver disease includes a wide range of pathology ranging from accumulation of fat only (isolated steatosis), to accumulation of fat with associated inflammation and liver cell damage (hepatocyte ballooning), collectively termed as metabolic dysfunction associated steatohepatitis (MASH; previously known as non-alcoholic steatohepatitis; NASH), and increasing degrees of fibrosis up to cirrhosis (F0-4).[3] Worsening stages of the disease, from isolated steatosis to MASH with fibrosis to cirrhosis, are associated with progressively increased risk of adverse clinical outcomes.[4]

There are currently no approved pharmacotherapies for MASLD and MASH. Given the high prevalence of these conditions, there is a very active field of clinical trials evaluating treatments. Histological classification is needed at baseline in phase 2b and 3 trials for the identification of MASH and staging of fibrosis to identify eligible patients for the studies. However, substantial interobserver variation has been described in how histological disease severity is assessed, even amongst expert pathologists.[5,6] The need for biopsy and the inherent limitations in its interpretation present major challenges in trial conduct as the recruitment process can be associated with high screen failure rates, when patients undergo biopsy but their disease severity is outside the trial inclusion criteria.[7] There is therefore a need for non-invasive screening strategies that can identify those who are more likely to meet histology eligibility criteria.[8] Regulators would consider treatments for patients with MASH+F2-3 or MASH+F4, and screening strategies have been described in the literature that have examined a at-risk MASH (MASH+F2-4)[9–13] encompassing both of the regulatory target conditions. To put our work in the context of the literature, we evaluated the performance of non-invasive tests (NITs) to screen for MASH+F2-4, but we also report how NITs perform if used to screen for MASH+F2-3 and MASH+F4, as these two target conditions have not been examined before.

Key points

Metabolic dysfunction-associated steatohepatitis (MASH) involves the simultaneous presence of fat, inflammation and scarring in the liver, affecting about one in 10 adults. Trials of several drugs targeting this disease are currently underway, and others will be started in future. However, participants in these trials are currently screened for eligibility by costly and invasive sampling of the liver tissue, a process called biopsy. In this work, we evaluated non-invasive tests applied for the screening of suitable participants for MASH trials. We have found that liver stiffness measurements combined with blood-based markers could reduce the proportion of potential participants by up to 80%. Such performance could decrease the costs related to running clinical trials and thus accelerate drug development.

Non-invasive tests have been extensively studied as risk stratification tools in clinical practice, where the main aim is to avoid liver biopsies where possible, mainly in patients with low NIT scores who have a low likelihood of clinically significant liver fibrosis.[14] However, data on the application of simple NITs as part of screening strategies for selecting patients for biopsy before inclusion in trials are limited compared with studies examining validation of diagnostic performance in routine clinical practice.

We previously conducted a large individual participant data meta-analysis (IPDMA) for the diagnostic accuracy of widely available NITs (liver stiffness measurement by vibration controlled transient elastography [LSM-VCTE], FIB-4 and NAFLD Fibrosis score [NFS]) for advanced fibrosis (F3-4).[15] The aim of this work was to use the IPDMA data set and evaluate whether the FibroScan-AST (FAST) score and these widely available NITs could be used as part of

screening strategies aimed to reduce the screen failure rate in clinical trials. We aimed to evaluate the performance of both single NITs and sequential application of NITs.

2 | METHODS

This IPDMA report was prepared in accordance with the recommendations of the PRISMA-IPD Statement.[16] The project was registered as PROSPERO CRD42019157661. Details of the selection criteria of studies, participants, index tests and reference standard, as well as details of quality and bias assessment of studies, establishing collaborations and data verification methods have been described previously.[15] Ethics approval was not sought for this meta-analysis as only anonymised data were provided by participating authors.

2.1 | Target conditions

MASH+F2-4 was the primary target condition of interest. MASH+F2-3, MASH+F4, significant fibrosis (F2-F4) and MASH were secondary target conditions. Metabolic dysfunction-associated steatotic liver disease was defined as NAS ≥4 with at least grade 1 in all three of lobular inflammation, ballooning and steatosis. The histological scoring used for this analysis was done as part of the primary studies by pathologists at the original study centres.

2.2 | Index tests and screening strategies

FAST score, LSM-VCTE, AST, FIB-4 and NFS were the evaluated index tests (Table S1). Single thresholds were used for screening MASH+F2-4 and MASH+F4, and a lower and upper threshold were applied to screen for MASH+F2-3. When screening for MASH+F2-3, the lower threshold was selected to be the same as the threshold with 90% sensitivity for F2-4. The upper threshold was selected based on its diagnostic accuracy for F4.[6] When screening for MASH+F2-3 participants with NIT results either below the lower or above the upper threshold were considered test negatives. Only participants with test results between the two thresholds were test positives, and thus proceeding to liver biopsy. When evaluating sequential application of NITs to screen for MASH+F2-3, dual thresholds were used for both NITs. For the target conditions of MASH+F2-4, MASH+F4, significant fibrosis and MASH, we examined the performance of sequential application of NITs using a single cut-off.

2.3 | Statistical analysis

The individual participant data sets provided by the authors of the original studies were merged, a study identification variable was added, and descriptive statistical analysis of the data sets was conducted. Dichotomous variables were displayed as percentages, and continuous variables were reported as means with standard deviations, or medians with interquartile ranges according to the distribution of the data. Participants without information on individual NAS components, without CAP-VCTE data or without sufficient information to compute the FIB-4 and NFS scores were excluded from this analysis.

Analyses were performed per-protocol, as there was insufficient information on failed LSM-VCTE. Diagnostic performance was expressed as the area under the empirical receiver operating characteristic curve (AUROC) with 95% confidence intervals (95% CI), based on De Long's method. AUROCs were compared using De Long's test statistic.

For each screening strategy, we computed sensitivity, specificity, number and proportion of patients that would be selected for liver biopsy, screen failure rate (SFR, 1–positive predictive value), misclassification rate (proportion of false-negative and false-positive cases), number of true positives per 100 cases (TP_{100}), number of patients needing to be tested with NITs to identify one true-positive case ($NNT = 100/TP_{100}$). The relationship between NIT cut-offs, SFR and NNT was examined visually for each screening strategy. When screening for MASH+F2-4, false-negative results were classified into MASH+F2-3 or MASH+F4, and false-positive results were classified into MASH+F0-1, F0-1 without MASH or F2-4 without MASH and the relationship of these with NIT thresholds was examined visually.

For each screening strategy, the performance of cut-offs that achieved SFR of 50%, 33% and the minimum achievable SFR while maximising sensitivity are reported (see also Supporting Methods). In addition, when screening for MASH+F2-4, we evaluated sequential strategies with a FIB-4 cut-off of 1.3 in the first tier. This cut-off is already used in clinical practice as a gateway for referral to secondary care, so in reality, the populations with MASH/MASLD seen in many secondary care centres in Europe is preselected to have FIB-4 >1.3. In these analyses, FIB-4 was followed by either LSM-VCTE or FAST with an aim to reduce the proportion of patients needing LSM-VCTE, while maintaining the desired diagnostic performance. The performance of cut-offs maximising the Youden index (i.e., sensitivity+specificity-1), for 90% sensitivity, for 90% specificity and previously published cut-offs (Table S2) are also reported.

An alternative testing strategy was evaluated for the detection of MASH+F2-3 employing FAST as first test and LSM-VCTE as a second test, each with a single threshold. Participants with FAST less than the 90% sensitivity threshold to detect significant fibrosis were deemed to be at low risk and considered as test negatives, similarly to participants with LSM-VCTE greater than the 90% specificity threshold to detect cirrhosis. All other participants were deemed as test positives.

Subgroup analyses were conducted to explore whether the performance of NITs as screening tests was influenced by BMI (≥30 kg/m^2/<30 kg/m^2), the presence of T2DM, LSM-VCTE probe type (M or XL), age (≥65 years old/<65 years old) and biopsy length (≥20 mm/<20 mm).

All statistical analyses were performed using R[17] (version 4.2.3, R Foundation for Statistical Computing, Vienna, Austria) with the pROC[18] package; 95% confidence intervals were calculated using 500 stratified bootstrap replicates using the boot package.[19,20]

3 | RESULTS

3.1 | Study and population characteristics

The selection process, characteristics and quality of studies included in this IPDMA are detailed in Figure S1, Supporting Results and Table S3. Data were available from 8045 individual participants (termed the entire data set). Complete histology and LSM-VCTE and CAP-VCTE data were available in 2427 cases (termed the study data set). FAST, FIB-4 and NFS could be calculated in 2281 cases (termed the analysis data set), and these cases were included in this study. In the analysis data set, median age was 55 years, 1140 (50%) participants were female, 1118 (49%) had diabetes and 867 (38%) had BMI ≥30 kg/m^2. The prevalence of significant fibrosis (*F* ≥2) was 47% (*n* = 1073), of MASH was 51% (*n* = 1170), of MASH with at least significant fibrosis (MASH+F2-4) was 31% (*n* = 705), of MASH with F2 and F3 stages of fibrosis (MASH+F2-3) was 24% (*n* = 553) and of MASH with cirrhosis (MASH+F4) was 7% (*n* = 152). The prevalence of fibrosis stages was 20% for F0 (*n* = 456), 33% for F1 (*n* = 753), 18% for F2 (*n* = 411), 17% for F3 (*n* = 388) and 12% for F4 (*n* = 273). Other demographic, histology, serum test and NIT details are shown in Table 1. The demographics of the study data set (*n* = 2427) and the analysis data set (*n* = 2281) are shown in Table S4.

3.2 | Diagnostic performance of single non-invasive tests

3.2.1 | MASH+F2-4

Using a single cut-off value to rule out cases with more mild disease, the FAST score had an AUROC of .78, significantly higher than that of LSM-VCTE (.75; *p* = .001). A SFR = 34% could be achieved at a FAST threshold of .8 with a sensitivity of 26%, eight true-positive cases detected per 100 patients tested with FAST (NNT = 12) and 262/2281 (12%) patients would have been selected for screening biopsy. At an LSM-VCTE, threshold of 14.1 kPa sensitivity was 38%, SFR was 46%, 12 true-positive cases were detected per 100 patients tested with LSM-VCTE (NNT = 8) and 482/2281 (22%) patients would have been selected for screening biopsy. FIB-4 and NFS had corresponding AUROCs of .68 and .57 for identifying MASH+F2-4, both significantly lower than that of LSM-VCTE (p < .001). The minimum achievable screen failure rate was 42% for FIB-4 and 62% for NFS. The performance of other thresholds, including those chosen to fulfil prespecified performance criteria and thresholds previously reported in the literature, is detailed in Table S5.

The relationship between NIT cut-offs and screen failure rate and the number of patients who need to be tested for a single true-positive case is illustrated in Figure 1. The relationship of each category of false-positive (F0-1 with MASH, F0-1 without MASH and F2-4 without MASH) and false-negative (MASH+F2-3, MASH+F4) results with NIT thresholds is illustrated in Figure 2.

The performance of NITs for the detection of significant fibrosis (F2-4) or MASH is presented in Tables S6 and S7.

3.2.2 | MASH+F2-3

Applying two cut-offs to the same biomarker, a lower threshold to rule out mild disease and a higher threshold to exclude cases with cirrhosis, LSM-VCTE had an AUROC of .67 for the detection of MASH+F2-3. At a threshold combination of 5.9 and 28 kPa the sensitivity was 84%, SFR was 69%, 20 true positive cases were detected per 100 patients tested with LSM-VCTE (NNT = 5) and biopsy was needed in 1481/2281 (65%) patients. The performance of other thresholds is summarised in Table S8.

FAST, FIB-4 and NFS had corresponding AUROCs of .73, .60 and .58 for identifying MASH+F2-3 (Table S8). Threshold pairs of .7 and 4.63 for FIB-4 and -3.272 and 1.570 for NFS performed similarly to 5.9 and 28 kPa for LSM-VCTE. The literature-based threshold pair of .35 and .67 for FAST, while having low sensitivity, performed similarly to the other tests in terms of SFR and yielded the lowest proportion of participants needing a liver biopsy. The relationship between NIT cut-offs and screen failure rate and number of patients who need to be tested for a single true-positive case is illustrated in Figure S3.

A second testing approach employing a single FAST threshold to rule out mild disease and a single LSM-VCTE threshold to rule out advanced disease had a sensitivity of 72%, screen failure rate of 65%, identified 17 true positive cases per 100 patients tested with NITs and led to biopsy in 1123/2281 (49%) of patients tested (Table S8).

3.2.3 | MASH+F4

Using a single cut-off value to rule out cases without MASH with cirrhosis, LSM-VCTE had an AUROC of .84 for identifying MASH+F4 (Table S9). At a threshold of 10 kPa sensitivity was 90%, SFR was 84%, six true-positive cases were detected per 100 patients tested with LSM-VCTE (NNT = 17) and biopsy was needed in 817/2281 (36%) patients. The performance of other thresholds including those chosen to fulfil prespecified performance criteria are detailed in Table S9.

FAST, FIB-4 and NFS had corresponding AUROCs of .79, .81 and .76 for identifying MASH+F4 (Table S9). Thresholds of .47 for FAST, 1.15 for FIB-4 and -1.866 for NFS performed similarly to 10 kPa for LSM-VCTE; however, more patients would need liver biopsy if screened using FAST, FIB-4 or NFS.

TABLE 1 Demographic details of the subgroup of patients in whom LSM-VCTE, FIB-4 and NFS were available.

	Analysis data set ($n=2281$)	F2-4 ($n=1073$)	MASH ($n=1170$)	MASH+F2-4 ($n=705$)	MASH+F2-3 ($n=553$)	MASH+F4 ($n=152$)
Females (%)	50	53	53	56	53	66
BMI ≥30 kg/m^2 (%)	38	43	45	47	50	36
Diabetes (%)	49	60	52	58	58	57
Age (years)[a]	55 (21)	58 (18)	55 (21)	57 (18)	57 (19)	62 (16)
BMI (kg/m^2)[a]	29 (7)	29 (7)	30 (7)	30 (7)	30 (7)	29 (8)
Biopsy data						
Steatosis (%)						
S0/S1/S2/S3	11/35/33/21	7/37/33/23	0/19/49/32	0/25/45/30	0/22/46/32	0/32/44/24
Ballooning (%)						
B0/B1/B2	32/49/19	16/53/31	0/64/36	0/54/46	0/57/43	0/43/57
Inflammation (%)						
I0/I1/I2/I3	19/56/22/3	12/53/30/5	0/56/38/6	0/49/42/9	0/51/42/7	0/42/45/13
Fibrosis (%)						
F0/F1/F2/F3/F4	20/33/18/17/12	0/0/38/37/25	7/33/22/25/13	0/0/37/41/22	0/0/47/53/0	0/0/0/0/100
NAS score[a]	4 (2)	4 (2)	5 (2)	5 (2)	5 (2)	5 (2)
MASH (%)	51	66	100	100	100	100
Liver function tests						
ALT (IU/L)[a]	50 (52)	54 (55)	64 (63)	65 (60)	69 (65)	53 (51)
AST (IU/L)[a]	39 (32)	46 (40)	48 (41)	53 (43)	53 (43)	54 (39)
Platelets (×10^9/L)[b]	230 (75)	209 (73)	230 (77)	214 (73)	227 (70)	168 (67)
Albumin (g/L)[b]	24 (20)	25 (20)	26 (20)	27 (19)	43 (5)	41 (5)
GGT (IU/L)[a]	55 (68)	68 (81)	62 (72)	76 (82)	75 (84)	87 (78)
Non-invasive tests						
LSM (kPa)[a]	7.9 (7.2)	11.5 (9.4)	9.3 (8.0)	11.6 (9.1)	10.3 (7.3)	17.7 (13.7)
FIB-4[a]	1.3 (1.2)	1.8 (1.7)	1.4 (1.3)	1.8 (1.7)	1.6 (1.3)	3.0 (2.6)
NFS	−.1 (2.2)	.4 (3.1)	−.2 (2.7)	.1 (2.9)	−.9 (1.9)	.1 (1.9)
FAST[a]	.4 (.4)	.6 (.4)	.6 (.4)	.7 (.3)	.6 (.3)	.7 (.2)

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body-mass index; FIB-4, fibrosis 4 score; GGT, gamma-glutamyl transferase; LSM, liver stiffness measurement; NAS, NAFLD activity score; NFS, NAFLD fibrosis score.

[a]Data are reported as median (IQR).

[b]Data are reported as mean (SD).

The relationship between NIT cut-offs and screen failure rate and number of patients who need to be tested for a single true-positive case is illustrated in Figure S4.

3.3 | Diagnostic performance of sequential combinations of non-invasive tests

3.3.1 | MASH+F2-4

FIB-4 followed by LSM-VCTE, FIB-4 followed by FAST, and NFS followed by FAST achieved screen failure rates of 33% or lower when screening for MASH+F2-4 at the expense of being able to detect only a few true-positive cases of 100 patients tested: respectively 6, 8 and 8 patients corresponding to threshold combinations of FIB-4: 2.00 and LSM-VCTE: 15.5 kPa, FIB-4: 1.40 and FAST: .65, and NFS: −.150 and FAST: .70 (Table S10).

Other cut-offs were also evaluated: thresholds corresponding to 90% sensitivity for a single NIT (FIB-4 at a threshold of .80 followed by LSM-VCTE threshold of 6.2 kPa) had a sensitivity of 81%, SFR of 54%, identified 25 true-positive cases per 100 patients tested with NITs and led to biopsy in 54% of patients tested), and thresholds corresponding to 90% sensitivity or 50% SFR for a single NIT (FIB-4 at a threshold of .8 followed by LSM-VCTE 8.4 kPa had a sensitivity of 68%, SFR of 48%, identified 21 true-positive cases per 100 patients tested with NITs and led to screening biopsy in 40% of patients tested with NITs (Table S10).

Figure 3 shows the relationship between FIB-4 and NFS thresholds combined with LSM-VCTE (Figure 3A–D) and FAST (Figure 3E–H) thresholds and screen failure rate and the number of patients

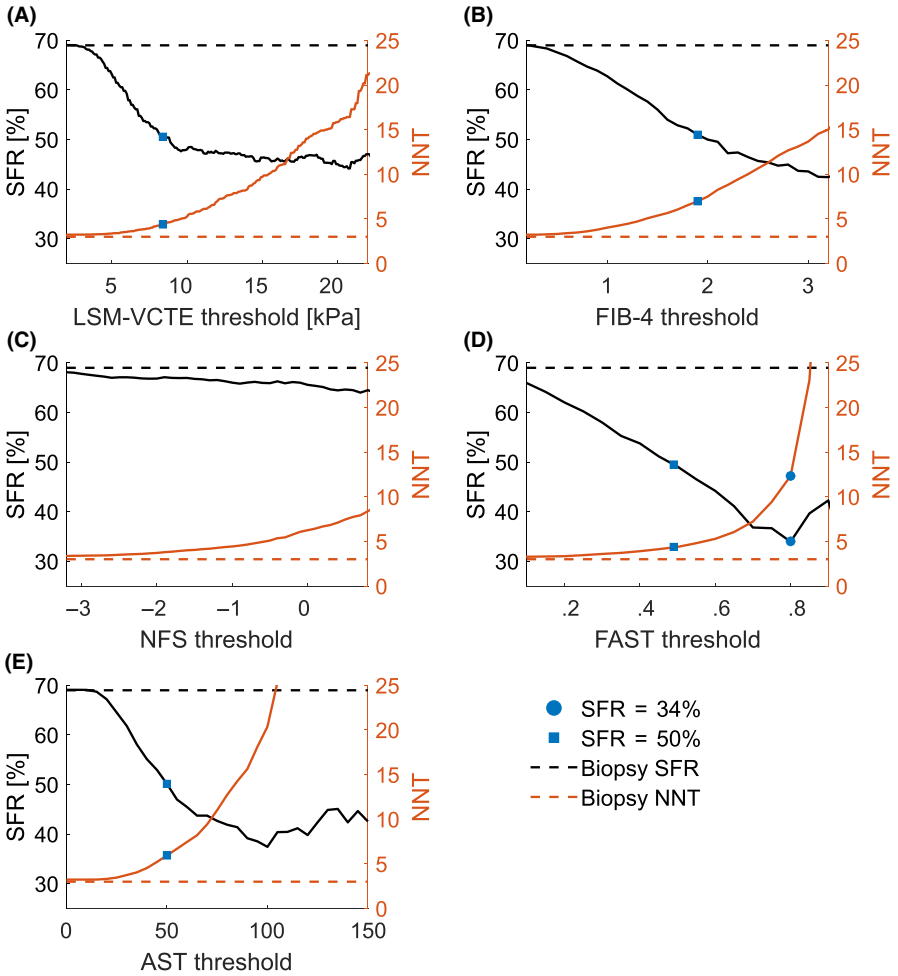

FIGURE 1 Screen failure rate and number of patients who need to be tested to identify a single true-positive case over the possible threshold ranges of (A) liver stiffness measurement by vibration controlled transient elastography (LSM-VCTE), (B) Fibrosis-4 Index (FIB-4), (C) NAFLD Fibrosis Score (NFS), (D) FibroScan-AST (FAST) score, and (E) AST when used in single test screening strategies for MASH+F2-4. Markers represent performance parameters at SFR of 34% and 50%.

who need to be tested for a single true-positive case for NIT combinations.

The number of true positive cases identified per 100 patients tested dropped with decreasing SFR due to an increased number of false-negative cases. The performance of all screening strategies evaluated at thresholds that achieved SFR of 33% and 50% is summarised in Figure 4.

A more detailed comparison of diagnostic performance of single NITs and sequential combinations of NITs at thresholds providing SFR = 33%, as well as for a literature-defined cut-off widely accepted by the community is presented in Table 2.

3.3.2 | MASH+F2-3

A FIB-4 threshold pair of .7 and 4.63 followed by an LSM-VCTE threshold pair of 5.9 and 28 kPa had a sensitivity of 66%, SFR of 67%, identified 16 true positive cases per 100 patients tested with NITs and led to biopsy in 1094/2281 (48%) of patients tested. An NFS threshold pair of −3.272 and 1.766 followed by an LSM-VCTE threshold pair of 5.9 and 28 kPa had a sensitivity of 75%, SFR of 67%, identified 18 true-positive cases per 100 patients tested with NITs and led to biopsy in 1257/2281 (55%) of patients tested with NITs. A FIB-4 threshold pair of .7 and 4.63 followed by a FAST threshold pair of .35 and .67 had a sensitivity of 35%, SFR of 70%, identified eight true-positive cases per 100 patients and led to biopsy in 603/2281 (26%) of patients tested with NITs. Finally, a NFS threshold pair of −3.272 and 1.766 followed by a FAST threshold pair of .35 and .67 had sensitivity of 34%, SFR of 69%, identified 8 true positive cases per 100 patients and led to biopsy in 632/2281 (28%) of patients tested with NITs (Table S11).

FIB-4 and NFS followed by either LSM-VCTE or FAST achieved only screen failure rates above 50% when screening for MASH+F2-3. In exchange, FIB-4+LSM-VCTE and NFS+LSM-VCTE NIT combinations were able to identify a single true positive case by only testing 6 patients for most threshold pairs (Table S11). Figure S5 shows the relationship between NIT thresholds and screen failure rate and number of patients who need to be tested for a single true positive case for NIT combinations where FIB-4 and NFS are applied first.

3.3.3 | MASH+F4

FIB-4 followed by LSM-VCTE had a sensitivity of 59%, SFR of 76%, identified four true-positive cases per 100 patients tested with NITs (NNT = 26) and led to biopsy in 365/2281 (16%) of patients tested for a threshold combination of .7 and 16 kPa (Table S12). The NAFLD fibrosis score followed by LSM-VCTE for thresholds

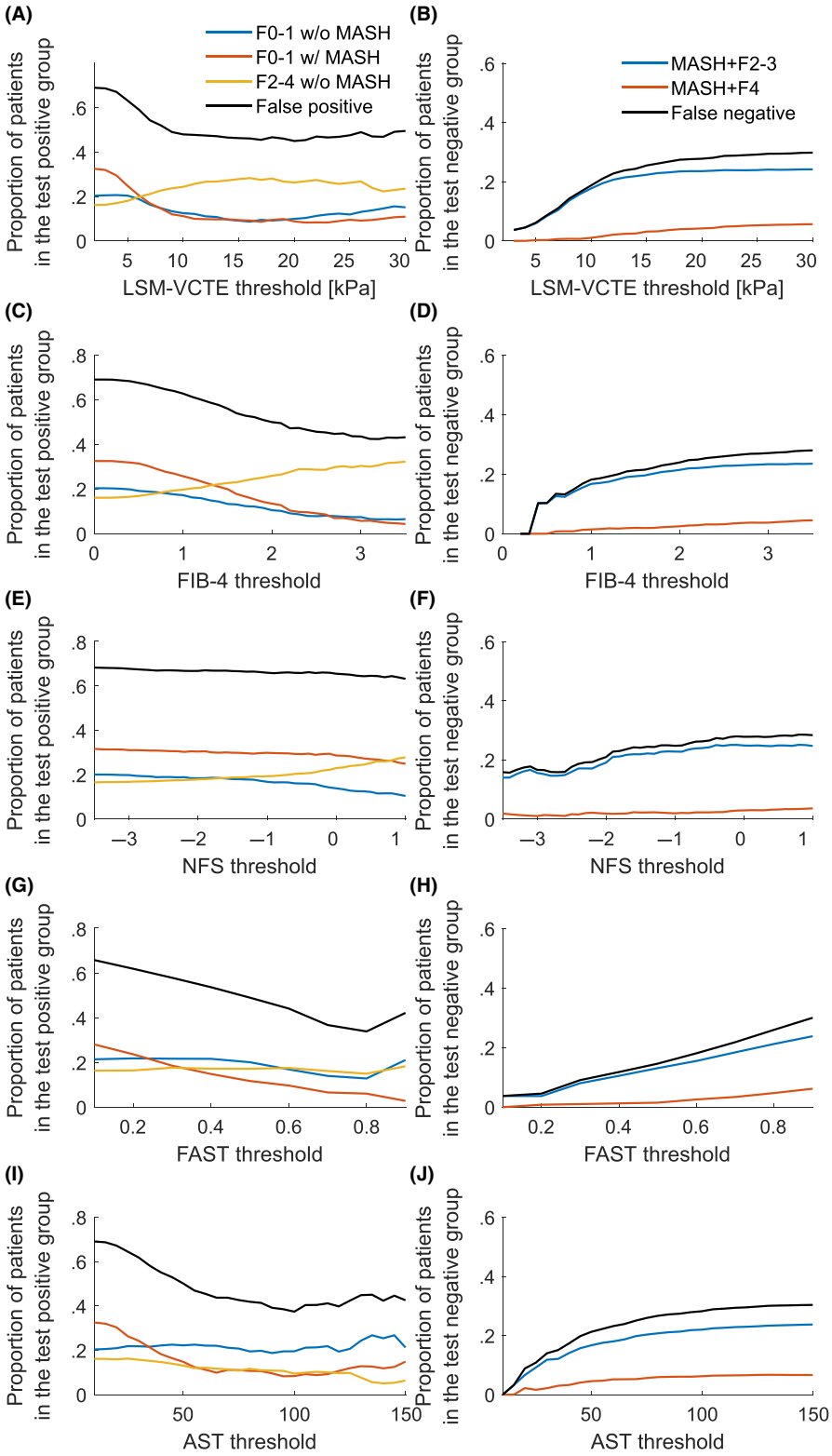

FIGURE 2 False-positive (A, C, E, G, and I) and false-negative (B, D, F, H, and J) cases classified into F0-1 with MASH, F0-1 without MASH and F2-4 without MASH (for false positives), and into MASH+F2-3 and MASH+F4 (for false negatives) subcategories for each NIT.

of −3.272 and 16 kPa yielded a similar performance: sensitivity of 59%, SFR of 76%, identified four true-positive cases per 100 patients tested with NITs and 372/2281 (16%) of patients biopsied. FIB-4 followed by FAST with thresholds of .7 and .35 had 91% sensitivity, 89% SFR, identified six true-positive cases per 100 patients tested with NITs and led to biopsy in 1212/2281 (53) of patients. The NAFLD fibrosis score followed by FAST with thresholds of −3.272 and .35 performed similarly to FIB-4 followed by FAST with thresholds of .7 and .35. The performance of other threshold combinations is detailed in Table S12. The relationship between NIT cut-offs and screen failure rate and number of patients who need to be tested for a single true-positive case is illustrated in Figure S6.

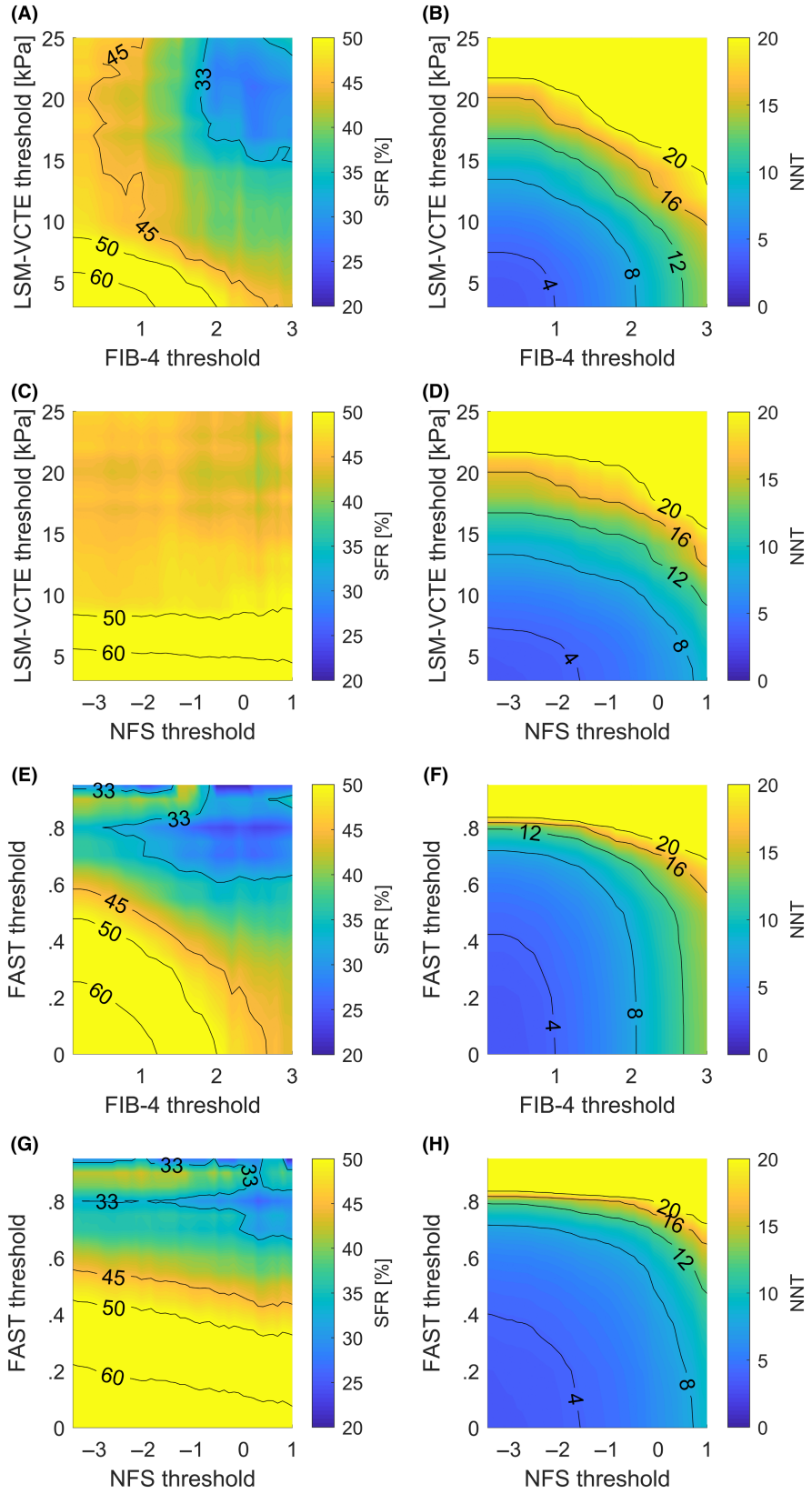

FIGURE 3 Screen failure rate and number of patients who need to be tested to identify a single true-positive case over the possible threshold ranges of fibrosis-4 score (FIB-4), NAFLD fibrosis score (NFS), liver stiffness measurement by vibration controlled transient elastography and FibroScan-AST score (FAST) when LSM-VCTE is applied after (A) FIB-4 or (B) (NFS) and when FAST is applied after (C) FIB-4 or (D) NFS, in sequential combinations of tests in screening strategies for MASH+F2-4.

3.3.4 | Subgroup analysis

When considering the detection of patients with MASH+F2-4, LSM-VCTE performed significantly better in the group of patients who had BMI < 30 kg/m^2 (27%), had no T2DM (26%), were scanned using an M probe (28%) or were younger than 65 years (Table S13). In participants where results from both the M and XL probes were available ($n = 180$), the diagnostic performance did not differ between the

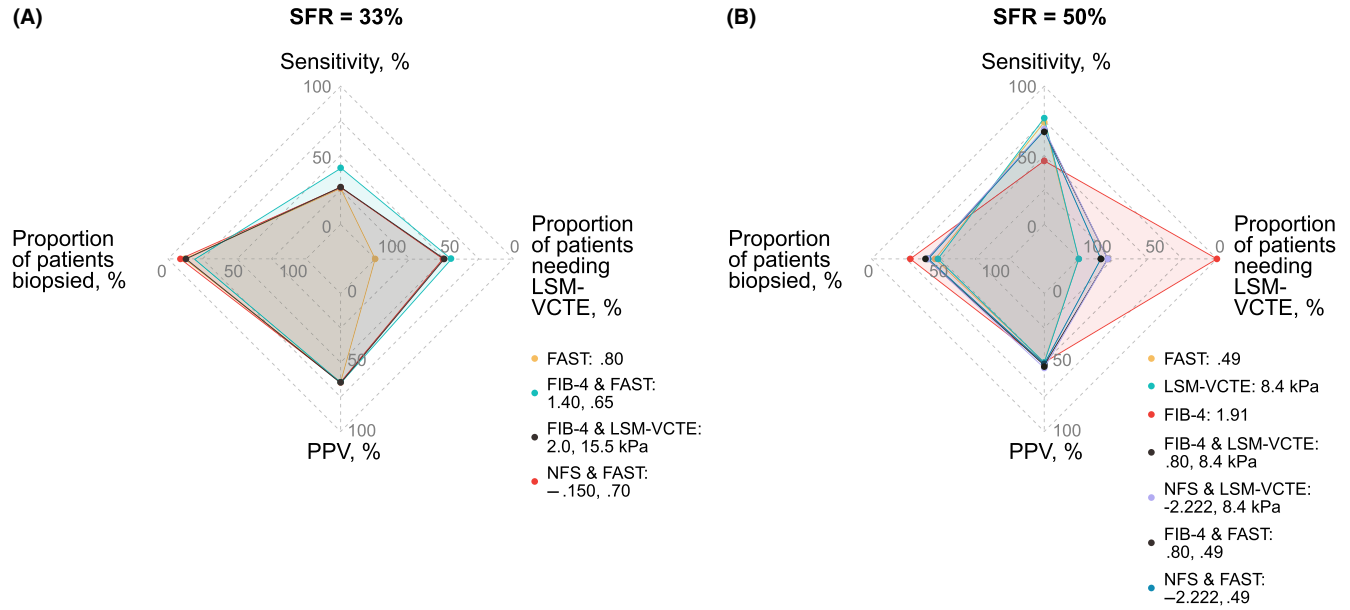

FIGURE 4 Radar charts comparing the diagnostic performance of single non-invasive tests (NITs) and the sequential combination of FIB-4 and NFS with LSM-VCTE and FAST for two thresholds corresponding to screen failure rate (SFR) of 33% (A) and 50% (B). Performance is described using four variables: sensitivity, positive predictive value (PPV), proportion of patients biopsied, and proportion of patients needing LSM-VCTE. Sensitivity and PPV scales start at 0% from the centre of each radar chart, while the scale for proportion of patients biopsied and the proportion of patients needing LSM-VCTE start at 0% from the outer vertices of each radar chart. Rectangles with larger area represent more optimal diagnostic performance. Proportion of patients biopsied, proportion of patients needing LSM-VCTE, and PPV are presented for a constant proportion of patients with MASH+F2-4 of 31%.

two probes (AUROC .75 vs .72, $p = .273$). The NAFLD fibrosis score had a significantly better performance in the subgroup of participants without T2DM, but there was no difference in its performance between BMI and age subgroups (Table S13). The presence of T2DM and age did not modify diagnostic performance of FIB-4 (Table S13). The FAST score performed significantly better in participants with BMI < 30 kg/m^2 and in patients without T2DM (Table S13).

FIB-4 alone was impacted by biopsy length, providing higher AUROC for patients with shorter than 20 mm samples (Table S13).

The performance of LSM-VCTE was impacted by the presence of T2DM, obesity and age >65 years and these three factors seemed to affect performance in a stepwise manner (i.e., presence of all three characteristics affected performance more than when only one of these factors was present). Other NITs did not appear to be affected in a similar manner (See Table S13 for details).

4 | DISCUSSION

In this large individual participant data meta-analysis, we examined how NITs could be used to screen participants and identify those likely to have more advanced disease and meet the inclusion criteria of a clinical trial. We chose MASH+F2-4 as the main target condition in our study as this is most relevant to clinical sites. This is for two reasons. First, in practical terms, clinical sites may be recruiting to both cirrhotic and non-cirrhotic MASH trials removing the need to screen separately for MASH+F2-3 and MASH+F4. Second, even if a clinical site is looking to screen for enrolment in a non-cirrhotic MASH trial, diagnosis of cirrhosis can still inform on the medical management of the patients at the clinical site. Therefore, even if cases with MASH+F4 would be screen failures from the study enrolment view point, this diagnosis is still worthwhile and not a screen failure from the clinical site's perspective. For the benefit of completeness, we also examined the performance of NITs to screen for MASH+F2-3 and MASH+F4 separately.

The proportion of participants with MASH+F2-4 was 31% in our study group, translating to a 69% screen failure rate if all of them were to undergo liver biopsy before inclusion in a clinical trial. We examined strategies to reduce the screen failure rates while also evaluating the effects of threshold choice on a range of parameters, including sensitivity and the number of patients who would be selected for biopsy. To account for the increasing use of simple NITs such as FIB-4 in community screening settings, we also specifically examined FIB-4 threshold of 1.3 as the first tier of screening in sequential approaches. Our main finding is that sequential applications of simple NITs (FIB-4 or NFS) followed by LSM-VCTE or FAST achieve similar diagnostic performance to LSM-VCTE or FAST alone as screening tests for MASH+F2-4. This has the advantage of needing fewer LSM-VCTE scans, which would have favourable cost implications in some settings.

If the NITs we examined are used exclusively as screening tools for MASH+F2-3 or MASH+F4, then only small reductions in screen failure rates can be achieved compared to having no screening at all, but with the benefit of needing to perform fewer biopsies. For

TABLE 2 Diagnostic performance of screening strategies for MASH+F2-4 at thresholds for achieving screen failure rate of 33%, or the lowest achievable screen failure rate.

	AUC	Thresholds	Sens, %	Spec, %	Pts biopsied, *n* (%)	SFR, %	Misclass, %	TP_{100}	NNT[a]	Proportion of patients needing LSM-VCTE, *n* (%)
Biopsy	–	–	–	–	2281 (100)	69	0	31	3	0 (0)
Newly derived thresholds										
FAST	.78 (.76–.80)	.80	26 (23–29)	94 (93–95)	262 (11)	34 (29–41)	27 (25–29)	8 (7–9)	12 (11–15)	2281 (100)
LSM-VCTE	.75 (.73–.77)	14.1 kPa	38 (35–42)	85 (83–87)	482 (21)	46 (42–51)	29 (27–31)	12 (10–13)	8 (8–10)	2281 (100)
AST	.71 (.69–.74)	100	16 (13–19)	96 (95–97)	171 (7)	37 (29–45)	29 (27–31)	5 (4–6)	20 (17–25)	0 (0)
FIB-4 followed by LSM-VCTE	–	2.00 15.5 kPa	20 (17–22)	96 (94–97)	209 (9)	33 (28–40)	28 (26–30)	6 (5–7)	16 (14–20)	608 (27)
FIB-4 followed by FAST	–	1.40 .65	40 (36–45)	91 (90–92)	428 (19)	34 (30–39)	25 (23–27)	12 (11–14)	8 (7–9)	1063 (47)
NFS followed by FAST	–	−.150 .70	27 (24–30)	94 (93–95)	288 (13)	33 (27–39)	27 (25–28)	8 (7–10)	12 (10–15)	1138 (50)
Literature-based thresholds										
FIB-4 followed by LSM-VCTE	–	1.30 8.0 kPa	56 (42–61)	79 (77–81)	723 (32)	46 (43–50)	28 (27–30)	17 (16–19)	6 (5–6)	1175 (52)
		1.30 10.0 kPa	47 (44–52)	84 (82–85)	580 (25)	44 (40–48)	28 (26–29)	15 (13–16)	7 (6–8)	1175 (52)
FIB-4 followed by FAST	–	1.30 .35	64 (60–68)	74 (72–76)	865 (38)	48 (45–52)	30 (28–31)	20 (18–22)	5 (5–6)	1175 (52)
		1.30 .67	39 (36–44)	91 (90–92)	419 (18)	34 (29–38)	25 (23–27)	12 (11–14)	8 (7–9)	1175 (52)

Note: 95% confidence intervals were estimated with 500 bootstrap replicates.

Abbreviations: AUC, area under the curve; Misclass, misclassified; NNT, number of patients that need to be tested test to identify one true-positive case per 100 patients; Pts, patients; Sens, sensitivity; SFR, screen failure rate; Spec, specificity; TP_{100}, true positives identified from a hypothetical group of 100 patients.

[a]NNT reported to the nearest whole number.

example, the prevalence of MASH+F2-3 in our study was 24%, meaning that SFR would be 76% if no screening was applied. Using FAST cut-offs of .35 and .67 to select patients for biopsy can reduce SFR to 72% and avoid biopsy in 65% of cases. Likewise, without screening, that is, 100% of patients undergo liver biopsy, the SFR for MASH+F4 is 93%, but using an LSM-VCTE threshold of >10 kPa to select patients for biopsy reduces SFR to 84% with only 36% of patients needing to undergo liver biopsy.

Our work should be considered in the context of other studies that have derived scores for the diagnosis of at-risk MASH. The study describing the FAST score is particularly relevant, as the participant groups used to validate the FAST score overlap with some of the datasets in our meta-analysis. While the work by Newsome et al.[9] did not evaluate the sequential application of NITs, its results do strengthen the argument that the inclusion of hepatic lipid content measurements are needed to improve diagnostic performance[9,10] and our results echo the superiority of the FAST score over LSM-VCTE alone.

The combination of FIB-4 with liver stiffness measurements by magnetic resonance elastography (MRE) has been described by the MEFIB score.[21] However, the MEFIB score has only been evaluated at this time using the concurrent measurement of both LSM by magnetic resonance elastography (LSM-MRE) and FIB-4 in all patients, rather than the sequential approach we use. While MEFIB demonstrated very high positive predictive value (>90%), it has only been examined for the identification of significant fibrosis (F2-4) alone, and not for MASH with significant fibrosis (MASH+F2-4), the latter being more relevant for clinical trial enrolment and was therefore assessed in our study. Beyond MEFIB, a number of studies have shown that adding hepatic fat content, using MRI proton density fat fraction (MRI-PDFF) or additional blood-based markers (AST, ALT), can increase diagnostic accuracy of MRE for detecting MASH with fibrosis.[10,22,23]

Other studies of combined markers have also shown promising performance[11,12,24,25]; however, these combinations include NITs that may not be widely available in clinical practice or are highly resource-intensive (e.g., NIS4[12] comprises miR-34a-5p, alpha-2 macroglobulin, YKL-40, and glycated haemoglobin; NIS2+[26] comprises miR-34a-5p, YKL-40, MAST[10] comprises magnetic resonance based proton density fat fraction [MRI-PDFF], magnetic resonance elastography [MRE], and AST; cTAG[11] uses magnetic resonance imaging to measure cT_1, AST and fasting serum glucose).

We focussed our analyses on the performance of NITs as screening tests for MASH+F2-4 where patients with NIT values above a certain threshold are selected for biopsy. In contrast to our analyses, other studies in the literature[9,10,12] also examine the performance of NITs using dual cut-offs where only patients in the indeterminate zone are selected for biopsy to determine the disease stage. The dual cut-off approach may be more relevant for application in routine practice rather than in the clinical trial screening, where it is important to identify the most participants having the target condition with the lowest possible screen failure rate. The identification of MASH with at least significant fibrosis will also be an important aspect in developing models of care in MASLD.[27,28]

We provide details of the diagnostic performance of various thresholds for various screening strategies, and we demonstrate the trade-off between screen failure rates and the number of true-positive cases identified. Thresholds that can achieve low screen failure rate would seem appealing, but they are not necessarily optimal as they necessitate more patients to be screened. This in turn could lead to the need for more trial sites to meet the recruitment target. Therefore, health economic and logistical considerations would need to be considered in determining optimal trial screening thresholds. Furthermore, NIT thresholds with high sensitivity for MASH+F2-4 are lower than the established rule-out thresholds for advanced fibrosis that are being used in primary referral pathways,[29,30] suggesting that in current practice a proportion of patients with at-risk MASH who are at risk of disease progression may not be identified.

Our subgroup analyses suggest superior diagnostic performance of LSM-VCTE in patients with BMI < 30 kg/m^2, patients without T2DM and those who had their measurement performed with an M probe. These findings should be interpreted carefully, as our data sets were collected between 2003 and 2017, during which period LSM-VCTE has seen several upgrades: the introduction of the XL probe, followed by the support for the Automated Probe Selection tool. This evolution meant that earlier studies may have collected data using the inappropriate probe, introducing BMI-related bias in measurements. The dependence of performance on the presence of T2DM may also be explained by collinearity with BMI.

In this work, we also report that markers previously extensively validated for the diagnosis of significant and advanced fibrosis (LSM-VCTE, FIB-4, and NFS) only have moderate diagnostic performance when considered screening for MASH+F2-4. This is in keeping with previous studies that report a similar performance for diagnosing MASH.[23]

It should be noted that individual NITs as well as their sequential combinations were compared against an imperfect reference standard of liver biopsy which has sampling-, and inter- and intra-observer variability.[5,6,31] This has been shown to lead to the underestimation of diagnostic performance of NITs[32] and in our study may have been further compromised by centre-level bias due to local biopsy reporting. In contrast, there is a move to consensus pathology reporting in clinical trials, which can also improve screen failure rates.

A limitation of our study is that we only had one data set from Australia and the USA, thus limiting the available range of patient BMI and making it difficult to generalise our findings to a global setting. As a lot of clinical trials are carried out in the USA, further validation in populations from this territory would be required. Some of the data in our study overlapped with data in the study that derived the FAST score,[9] which may have introduced bias to our analysis. However, while the study describing the FAST score examined the application of this score in isolation, we examine the serial application with other simple tests. Lastly, the changing practice and use of NIT screening strategies in the community or before selecting patients for biopsy could not be accounted for in this retrospective study with data collected over more than 10 years.

In summary, we have conducted a large individual participant data meta-analysis and shown that screening strategies using sequential application of simple NITs like FIB-4 or NFS followed by LSM-VCTE achieve similar diagnostic performance to LSM-VCTE alone. While this approach has been validated before in strategies to identify those at low risk of advanced fibrosis[15] in clinical practice, it has not been examined in screening strategies for clinical trials. This can have favourable cost implications by needing to perform fewer NITs in the second tier of testing, while at the same time reducing screen failure rates and the number of patients that need to undergo biopsy. Furthermore, screening with NITs to identify patients at high risk of MASH+F2-3 and MASH+F4 can lead to modest gains in SFR and sizeable reductions in the number of biopsies that need to be performed, a strategy that can carry favourable cost implications.

AUTHOR CONTRIBUTIONS

FEM, EAS, ANAJ, AG, TT, QMA, SAH, PMB and MP contributed to the planning and design of the study. JB, VL, CF, MLP, YY, TS, TO, SM, WKC, VWSW, TK, JW, S, ET, AL, DHL, AGH, AMD, ALM, HH, CA, MH, DHL, WK, AN, MY, PT, FL, AB, YPM, MN and ET collected and provided individual patient data. FEM, YV, JAL, PMB and MP performed statistical analyses and data interpretation. FEM and MP wrote the first draft of the manuscript. All co-authors approved the final version of the manuscript.

AFFILIATIONS

[1]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, OCMR, University of Oxford, Oxford, UK
[2]Department of Epidemiology and Data Science, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands
[3]Translational Gastroenterology Unit, University of Oxford, Oxford, UK
[4]Oxford NIHR Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust and the University of Oxford, Oxford, UK
[5]Laboratoire HIFIH, UPRES EA 3859, SFR ICAT 4208, Université d'Angers, Angers, France
[6]Service d'Hépato-Gastroentérologie et Oncologie Digestive, Centre Hospitalier Universitaire d'Angers, Angers, France
[7]Centre d'Investigation de la Fibrose Hépatique, Hôpital Haut-Lévêque, Bordeaux University Hospital, Pessac, France
[8]INSERM1312, Bordeaux University, Bordeaux, France
[9]Department of Medical Imaging, Iuliu Haţieganu University of Medicine and Pharmacy, Regional Institute of Gastroenterology and Hepatology "Prof.Dr. Octavian Fodor", Cluj-Napoca, Romania
[10]Department of Gastroenterology, School of Medicine, Marmara University, Istanbul, Turkey
[11]Department of Gastroenterology, School of Medicine, Recep Tayyip Erdoğan University, Rize, Turkey
[12]Gastroenterology and Hepatology Unit, Department of Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia
[13]Department of Oncology, Gastroenterology, Hepatology, Pulmonology and Infectious Diseases, University Hospital Leipzig, Leipzig, Germany
[14]Department of Gastroenterology, All India Institute of Medical Sciences, New Delhi, India
[15]Sheila Sherlock Liver Unit and UCL Institute for Liver and Digestive Health, Royal Free Hospital and University College London, London, UK
[16]Department of Translational Medicine and Surgery, Università Cattolica Del Sacro Cuore, Rome, Italy
[17]Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, Hong Kong
[18]Department of Internal Medicine, Gil Medical Center, Gachon University College of Medicine, Incheon, Korea
[19]Department of Vascular Medicine, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands
[20]Division of Liver and Pancreatic diseases, Department of Upper GI, Karolinska University Hospital, Stockholm, Sweden
[21]Department of Medicine, Huddinge, Karolinska Institutet, Stockholm, Sweden
[22]Department of Gastroenterology and Metabology, Ehime University Graduate School of Medicine, Touon, Ehime, Japan
[23]Division of Gastroenterology and Hepatology, Department of Internal Medicine, Seoul National University College of Medicine, Seoul Metropolitan Government Boramae Medical Center, Seoul, Republic of Korea
[24]Department of Gastroenterology and Hepatology, Saiseikai Suita Hospital, Suita, Japan
[25]Department of Gastroenterology and Hepatology, Yokohama City University School of Medicine, Yokohama, Japan
[26]Institute of Digestive Health and Liver Diseases, Mercy Medical Center, Baltimore, Maryland, USA
[27]Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA
[28]Department for Visceral Medicine and Surgery, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland
[29]Department of Biomedical Research, University of Bern, Bern, Switzerland
[30]Graduate School for Health Sciences (GHS), University of Bern, Bern, Switzerland
[31]Houston Research Institute, Houston Methodist Hospital, Houston, Texas, USA
[32]Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, California, USA
[33]Echosens, Paris, France
[34]Division of Hepatology, University Hospital Würzburg, Würzburg, Germany
[35]Internal Medicine Research Unit, Pfizer Inc, Cambridge, Massachusetts, USA
[36]Clinical Development and Operations, Global Product Development, Pfizer, Inc, Lake Mary, Florida, USA
[37]Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK
[38]Newcastle NIHR Biomedical Research Centre, Newcastle upon Tyne Hospitals NHS Trust, Newcastle upon Tyne, UK

ACKNOWLEDGEMENTS

LITMUS Investigators (see Supporting Information for list of LITMUS Investigators).

CONFLICT OF INTEREST STATEMENT

FEM, JAL, YV, EAS, JB, MLP, YY, AL, DHL, AGH, AMvD, ALM, CA, MH, DHL, TS, TO, AN, S, PJT, FL, AB, YPM, ET and **PMB** declare no conflict of interest. **ANAJ** and **MP** are shareholders of Perspectum, Oxford, UK. **SM** received honorarium fees from Echosens. **WKC** has served as a consultant or advisory board member for Roche, AbbVie, Boehringer Ingelheim and Novo Nordisk, and a speaker for Echosens, Hisky Medical and Viatris. **TK** and **JW** received unrestricted research grants from Echosens. **TK** has served as a speaker for Echosens. **ET** has served on the advisory boards for Pfizer, NovoNordisk, Orphalan and Alexion. **VWSW** has served as a consultant or advisory board member for 3V-BIO, AbbVie, Allergan, Boehringer Ingelheim, Center for Outcomes Research in Liver Diseases, Echosens, Gilead Sciences, Hanmi Pharmaceutical, Intercept, Merck, Novartis, Novo Nordisk, Perspectum, Pfizer, ProSciento, Sagimet Biosciences, TARGET PharmaSolutions, and Terns; and a speaker for AbbVie, Bristol-Myers Squibb, Echosens,

and Gilead Sciences. He has also received a research grant from Gilead Sciences for fatty liver research. **VL** reports consultancy for AbbVie, BMS, Echosens, Gilead Sciences, Intercept Pharmaceuticals, MSD, Myr-Pharma, Pfizer, Supersonic Imagine and Tillotts. **WK** has served as a speaker and consultant of Gilead, Boehringer-Ingelheim, Samil, Ildong, LG Chemistry, HK inno.N, GreenCross, Bukwang, Standigm, PharmaKing, KOBIOLABS, Eisai, Zydus, and Novonordisk, received grants from Gilead, Ildong, GreenCross, Bukwang, Pharmaking, Roche, Galmed, Novartis, Pfizer, Springbank, Altimmune, MSD, BMS, Dicerna, Enyo, and Hitachi-Aloka, and owns stocks in KOBIOLABS and Lepidyne. **MN** has been on the advisory board/consultant for 89BIO, Altimmune, BI, Gilead, cohBar, Cytodyn, Pfizer, GSK, Novo Nordisk, EchoSens, Madrigal, NorthSea, Prespecturm, Terns, Takeda, Sami-Sabina group, Siemens and Roche diagnostic; MN has received research support from Allergan, Akero, BMS, Gilead, Galmed, Galectin, Genfit, Conatus, Corcept, Enanta, Madrigal, Novartis, Pfizer, Shire, TERNS, Viking and Zydus; MN is a shareholder or has stocks in Anaetos, Chrownwell, Cytodyn, Ciema, Rivus Pharma and Viking. **MY** received research support from Kowa Co. Ltd. **CFP** is employed by Echosens, France. **AG** reports consultancy for AbbVie, Alexion, Bayer, BMS, CSL Behring, Eisai, Gilead, Heel, Intercept, Ipsen, Merz, MSD, Novartis, Pfizer, Roche, Sanofi-Aventis, Sequana; received research funding from Intercept, Falk, Novartis and was on the speakers bureau for AbbVie, Alexion, BMS, Burgerstein, CSL Behring, Falk Foundation, Gilead, Intercept, MSD, Merz, Novartis, Sequana. **TT**, and **CY** are employees of Pfizer, Inc. **QMA** is coordinator of the IMI2 LITMUS consortium and he reports research grant funding from Abbvie, Allergan/Tobira, AstraZeneca, GlaxoSmithKline, Glympse Bio, Novartis Pharma AG, Pfizer Ltd., Vertex; consultancy on behalf of Newcastle University for Abbott Laboratories, Acuitas Medical, Allergan/Tobira, Blade, BNN Cardio, Cirius, CymaBay, EcoR1, E3Bio, Eli Lilly & Company Ltd., Galmed, Genfit SA, Gilead, Grunthal, HistoIndex, Indalo, Imperial Innovations, Intercept Pharma Europe Ltd., Inventiva, IQVIA, Janssen, Kenes, Madrigal, MedImmune, Metacrine, NewGene, NGMBio, North Sea Therapeutics, Novartis, Novo Nordisk A/S, Pfizer Ltd., Poxel, ProSciento, Raptor Pharma, Servier, Viking Therapeutics; and speaker fees from Abbott Laboratories, Allergan/Tobira, BMS, Clinical Care Options, Falk, Fishawack, Genfit SA, Gilead, Integritas Communications, MedScape. **SAH** has research grants from Akero, Altimmune, Axcella-Cirius, CiVi Biopharma, Cymabay, Galectin, Genfit, Gilead Sciences, Hepion Pharmaceuticals, Hightide Therapeutics, Intercept, Madrigal, Metacrine, NGM Bio, Northsea Therapeutics, Novartis, Novo Nordisk, Poxel, Sagimet, Viking. He has received consulting fees from Akero, Altimmune, Alentis, Arrowhead, Axcella, Echosens, Enyo, Foresite Labs, Galectin, Genfit, Gilead Sciences, Hepion, HIghtide, HistoIndex, Intercept, Kowa, Madrigal, Metacrine, NeuroBo, NGM, Northsea, Novartis, Novo Nordisk, Poxel, Perspectum, Sagimet, Terns, and Viking. **HH**'s institutions have received research grants from Astra Zeneca, EchoSens, Gilead, Intercept, MSD, and Pfizer.

FUNDING INFORMATION

This individual participant data meta-analysis has been conducted as part of the imaging study in the LITMUS (Liver Investigation: Testing Marker Utility in Steatohepatitis) study. The LITMUS study is a large multicentre study aiming to evaluate biomarkers on Non-Alcoholic Fatty Liver Disease. The LITMUS study is funded by the Innovative Medicines Initiative 2 (IMI2) Joint Undertaking under Grant Agreement 777 377. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. The funder and the authors' institutions had no role in the development of the protocol for this study. This communication reflects the view of the LITMUS consortium and neither IMI nor the European Union and EFPIA are liable for any use that may be made of the information contained herein. Ferenc E. Mózes acknowledges support from a Sir Henry Dale Fellowship of the Wellcome Trust and the Royal Society [221 805/Z/20/Z].

ORCID

Ferenc E. Mózes https://orcid.org/0000-0002-1361-4349
Jenny A. Lee https://orcid.org/0000-0003-4024-0933
Yasaman Vali https://orcid.org/0000-0001-7002-118X
Arjun N. A. Jayaswal https://orcid.org/0000-0002-3272-2695
Victor de Lédinghen https://orcid.org/0000-0001-6414-1951
Monica Lupșor-Platon https://orcid.org/0000-0001-7918-1956
Yusuf Yilmaz https://orcid.org/0000-0003-4518-5283
Wah-Kheong Chan https://orcid.org/0000-0002-9105-5837
Thomas Karlas https://orcid.org/0000-0002-8109-8526
Johannes Wiegand https://orcid.org/0000-0001-9233-4064
Emmanouil Tsochatzis https://orcid.org/0000-0001-5069-2461
Vincent Wai-Sun Wong https://orcid.org/0000-0003-2215-9410
Dae Ho Lee https://orcid.org/0000-0002-8832-3052
Adriaan G. Holleboom https://orcid.org/0000-0002-2911-2917
Anne-Marieke van Dijk https://orcid.org/0000-0003-0831-527X
Hannes Hagström https://orcid.org/0000-0002-8474-1759
Won Kim https://orcid.org/0000-0002-2926-1007
Takeshi Okanoue https://orcid.org/0000-0002-2390-3400
Atsushi Nakajima https://orcid.org/0000-0002-6263-1436
Paul J. Thuluvath https://orcid.org/0000-0002-4374-4507
Annalisa Berzigotti https://orcid.org/0000-0003-4562-9016
Stephen A. Harrison https://orcid.org/0000-0001-8285-2204
Michael Pavlides https://orcid.org/0000-0001-9882-8874

REFERENCES

1. Rinella ME, Lazarus JV, Ratziu V, et al. A multi-society Delphi consensus statement on new fatty liver disease nomenclature. *Ann Hepatol*. 2023;29:101133. doi:10.1016/j.aohep.2023.101133
2. 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. doi:10.1038/nrgastro.2017.109
3. 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. doi:10.1002/hep.20701
4. Simon TG, Roelstraete B, Khalili H, Hagström H, Ludvigsson JF. Mortality in biopsy-confirmed nonalcoholic fatty liver disease:

results from a nationwide cohort. *Gut*. 2021;70(7):1375-1382. doi:10.1136/GUTJNL-2020-322786
5. Brunt EM, Clouston AD, Goodman Z, et al. Complexity of ballooned hepatocyte feature recognition: defining a training atlas for artificial intelligence-based imaging in NAFLD. *J Hepatol*. 2022;76:1030-1041. doi:10.1016/J.JHEP.2022.01.011
6. Davison BA, Harrison SA, Cotter G, et al. Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials. *J Hepatol*. 2020;73(6):1322-1332. doi:10.1016/j.jhep.2020.06.025
7. Friedman SL, Ratziu V, Harrison SA, et al. A randomized, placebo-controlled trial of cenicriviroc for treatment of nonalcoholic steatohepatitis with fibrosis. *Hepatology*. 2018;67(5):1754-1767. doi:10.1002/HEP.29477
8. Anstee QM, Castera L, Loomba R. Impact of non-invasive biomarkers on hepatology practice: past, present and future. *J Hepatol*. 2022;76(6):1362-1378. doi:10.1016/j.jhep.2022.03.026
9. Newsome PN, Sasso M, Deeks JJ, 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(4):362-373. doi:10.1016/S2468-1253(19)30383-8
10. Noureddin M, Truong E, Gornbein JA, et al. MRI-based (MAST) score accurately identifies patients with NASH and significant fibrosis. *J Hepatol*. 2021;76:781-787. doi:10.1016/J.JHEP.2021.11.012
11. Dennis A, Mouchti S, Kelly 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(1):1-11. doi:10.1038/s41598-020-71995-8
12. Harrison SA, Ratziu V, Boursier J, 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;5(11):970-985. doi:10.1016/S2468-1253(20)30252-1
13. Tamaki N, Imajo K, Sharpton S, et al. Magnetic resonance elastography plus Fibrosis-4 versus FibroScan–aspartate aminotransferase in detection of candidates for pharmacological treatment of NASH-related fibrosis. *Hepatology*. 2022;75(3):661-672. doi:10.1002/HEP.32145
14. Vali Y, Lee J, Boursier J, et al. Biomarkers for staging fibrosis and non-alcoholic steatohepatitis in non-alcoholic fatty liver disease (the LITMUS project): a comparative diagnostic accuracy study. *Lancet Gastroenterol Hepatol*. 2023;8:714-725. doi:10.1016/S2468-1253(23)00017-1
15. Mózes FE, Lee JA, Selvaraj EA, et al. Diagnostic accuracy of non-invasive tests for advanced fibrosis in patients with NAFLD: an individual patient data meta-analysis. *Gut*. 2022;71:1006-1019. doi:10.1136/gutjnl-2021-324243
16. Stewart LA, Clarke M, Rovers M, et al. Preferred reporting items for a systematic review and meta-analysis of individual participant data. *JAMA*. 2015;313(16):1657. doi:10.1001/jama.2015.3656
17. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2020.
18. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12(1):77. doi:10.1186/1471-2105-12-77
19. Canty A, Ripley B. boot: Bootstrap R (S-Plus) Functions. R package version 1.3-24. 2019.
20. Davison AC, Hinkley DV. *Bootstrap Methods and their Application*. Cambridge University Press; 1997. doi:10.1017/cbo9780511802843
21. Jung J, Loomba RR, Imajo K, et al. MRE combined with FIB-4 (MEFIB) index in detection of candidates for pharmacological treatment of NASH-related fibrosis. *Gut*. 2021;70:1946-1953. doi:10.1136/gutjnl-2020-322976
22. Dzyubak B, Li J, Chen J, et al. Automated analysis of multiparametric magnetic resonance imaging/magnetic resonance Elastography exams for prediction of nonalcoholic Steatohepatitis. *J Magn Reson Imaging*. 2021;54(1):122-131. doi:10.1002/JMRI.27549
23. Lee YS, Yoo YJ, Jung YK, et al. Multiparametric MR is a valuable modality for evaluating disease severity of nonalcoholic fatty liver disease. *Clin Transl Gastroenterol*. 2020;11(4):e00157. doi:10.14309/CTG.0000000000000157
24. Andersson A, Kelly M, Imajo K, et al. Clinical utility of MRI biomarkers for identifying NASH patients at high risk of progression: a multicenter pooled data and meta-analysis. *Clin Gastroenterol Hepatol*. 2022;20:2451-2461.e3. doi:10.1016/J.CGH.2021.09.041
25. Gao F, Huang J-F, Zheng KI, et al. Development and validation of a novel non-invasive test for diagnosing fibrotic non-alcoholic steatohepatitis in patients with biopsy-proven non-alcoholic fatty liver disease. *J Gastroenterol Hepatol*. 2020;35(10):1804-1812. doi:10.1111/JGH.15055
26. Harrison SA, Ratziu V, Magnanensi J, et al. NIS2+™, an optimisation of the blood-based biomarker NIS4® technology for the detection of at-risk NASH: a prospective derivation and validation study. *J Hepatol*. 2023;79(3):758-767. doi:10.1016/j.jhep.2023.04.031
27. Lazarus JV, Mark HE, Anstee QM, et al. Advancing the global public health agenda for NAFLD: a consensus statement. *Nat Rev Gastroenterol Hepatol*. 2022;19(1):60-78. doi:10.1038/s41575-021-00523-4
28. Lazarus JV, Anstee QM, Hagström H, et al. Defining comprehensive models of care for NAFLD. *Nat Rev Gastroenterol Hepatol*. 2021;18(10):717-729. doi:10.1038/s41575-021-00477-7
29. Srivastava A, Gailer R, Tanwar S, et al. Prospective evaluation of a primary care referral pathway for patients with non-alcoholic fatty liver disease. *J Hepatol*. 2019;71(2):371-378. doi:10.1016/j.jhep.2019.03.033
30. Moolla A, Motohashi K, Marjot T, et al. A multidisciplinary approach to the management of NAFLD is associated with improvement in markers of liver and cardio-metabolic health. *Frontline Gastroenterol*. 2019;10(4):337-346. doi:10.1136/flgastro-2018-101155
31. Ratziu V, Charlotte F, Heurtier A, et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology*. 2005;128(7):1898-1906. doi:10.1053/j.gastro.2005.03.084
32. Mehta SH, Lau B, Afdhal NH, Thomas DL. Exceeding the limits of liver histology markers. *J Hepatol*. 2009;50(1):36-41. doi:10.1016/J.JHEP.2008.07.039

SUPPORTING INFORMATION

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

How to cite this article: Mózes FE, Lee JA, Vali Y, et al. Diagnostic accuracy of non-invasive tests to screen for at-risk MASH—An individual participant data meta-analysis. *Liver Int*. 2024;00:1-14. doi:10.1111/liv.15914