Annals of Gastroenterology (2025) **38,** 681-690

Magnetic resonance elastography combined with fibrosis-4 index for diagnosing at-risk metabolic dysfunction-associated steatohepatitis: a systematic review and meta-analysis of diagnostic test accuracy studies

Konstantinos Malandris[a], Anastasia Katsoula[b], Tarek Nayfeh[c], Kalliopi Tsapa[d], Dimitra Tsapa[d], Georgios Kalopitas[e], Aris Liakos[a], Thomas Karagiannis[a], Eleni Theocharidou[a], Emmanouil Sinakos[f], Georgios Germanidis[e], Apostolos Tsapas[a,g]

Aristotle University of Thessaloniki, Greece; Georgetown University, Baltimore, Maryland, USA; Harris Manchester College, University of Oxford, UK

Abstract

Background Patients with metabolic dysfunction-associated steatohepatitis (MASH; nonalcoholic fatty liver disease activity score ≥4) and significant fibrosis (≥F2; at-risk MASH) are at increased risk for disease progression. Magnetic resonance elastography (MRE) combined with the fibrosis-4 (MEFIB) index enables the noninvasive diagnosis of at-risk MASH and significant fibrosis. We assessed the performance of the MEFIB index for ruling in/out both target conditions.

Methods We analyzed studies up to February 2025 assessing the performance of MEFIB index for ruling in (MRE≥3.3 kPa plus FIB-4≥1.6) and out (MRE<3.3 kPa plus FIB-4<1.6) at-risk MASH or significant fibrosis, using liver biopsy as the reference standard. We calculated pooled diagnostic accuracy estimates using bivariate random-effects models.

Results We included 7 studies with 3356 participants. For ruling in at-risk MASH, the MEFIB index yielded a pooled specificity of 0.94 (95% confidence interval [CI] 0.74-0.99), and a positive likelihood ratio (LRp) of 5.3 (95%CI 1.8-15.7). For ruling out at-risk MASH, the MEFIB index had a pooled sensitivity of 0.77 (95%CI 0.62-0.88) and a negative likelihood ratio (LRn) of 0.34 (95%CI 0.23-0.52). For ruling in significant fibrosis, the MEFIB index achieved a summary specificity of 0.93 (95%CI 0.85-0.97) with LRp 8.2 (95%CI 4.5-14.9). For excluding significant fibrosis, the pooled sensitivity and LRn of the MEFIB index were 0.88 (95%CI 0.79-0.94) and 0.16 (95%CI 0.08-0.31), respectively.

Conclusions MEFIB index has acceptable accuracy for diagnosing at-risk MASH and significant fibrosis. Proposed thresholds can be used to identify both target conditions in high prevalence settings and facilitate patient recruitment in clinical trials.

Keywords MEFIB index, metabolic dysfunction-associated steatohepatitis, fibrosis, systematic review, meta-analysis

Ann Gastroenterol 2025; 38 (6): 681-690

Introduction

Metabolic dysfunction-associated steatotic liver disease (MASLD) is characterized by the accumulation of hepatic fat in the presence of specific cardiometabolic risk factors, after the exclusion of secondary causes of liver steatosis [1,2]. Its progressive form, metabolic dysfunction-associated steatohepatitis (MASH), has emerged as the second most common indication for liver transplantation in the United States [3]. Individuals with MASH and significant fibrosis (F≥F2), referred to as "at-risk MASH", are at increased risk for disease progression and liver-related mortality, constituting the target population for MASH clinical trials [4].

Conflict of Interest: None

Correspondence to: Konstantinos Malandris, MD, MSc, Clinical Research and Evidence-Based Medicine Unit, Second Medical Department, Aristotle University of Thessaloniki, Konstantinoupoleos 49, 54642 Thessaloniki, Greece, e-mail: kostas_malandris@yahoo.gr

Received 24 June 2025; accepted 16 September 2025; published online 10 October 2025

DOI: https://doi.org/10.20524/aog.2025.1010

The requirement for specific histopathologic criteria to identify candidates for enrollment in MASH clinical trials raise significant challenges, one of which is the high rate of screening failure [4]. To mitigate this issue, and reduce the need for unnecessary liver biopsies, several noninvasive biomarkers have been proposed for the selection of potentially eligible participants [5]. Fibrosis-4 index (FIB-4) and vibration controlled transient elastography (VCTE) are the most validated biomarkers for the assessment of fibrosis, serving as initial steps of many recommended pathways [1,2,6]. However, their low positive predictive values (PPVs), attributed mainly to the low prevalence of MASLD with significant fibrosis, limit their ability to set the diagnosis [7].

Following the approval of resmetirom and semaglutide for MASLD, there is an even greater need to identify patients with at-risk MASH, ideally without requiring a liver biopsy [7]. In response, there has been a growing trend towards the development of sequential testing strategies that integrate serum-based and imaging-based indices [8]. Previous studies have shown that the combination of magnetic resonance elastography (MRE) and FIB-4 index, known as the MEFIB index, is superior to its individual components, and to the FibroScan-aspartate aminotransferase (FAST) score, for identifying candidates for MASH clinical trials [8,9]. We conducted a systematic review and meta-analysis to summarize, and critically appraise, findings from individual studies assessing the accuracy of the MEFIB index for diagnosing at-risk MASH and significant fibrosis.

Materials and methods

We conducted this systematic review and meta-analysis following a prespecified protocol registered in PROSPERO (CRD420251041430). Our methodology and results adhere to the Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) guidelines (Supplementary Table 1) [10].

Eligibility criteria

We included cross-sectional studies assessing the accuracy of the MEFIB index for diagnosing at-risk MASH or significant fibrosis (fibrosis stage ≥F2) in adults with MASLD, using liver biopsy as the reference standard. At-risk MASH was defined as MASH with nonalcoholic fatty liver disease activity score (NAS) ≥4 and fibrosis stage ≥F2. For the MEFIB index we considered only the diagnostic thresholds recommended by the respective American and European guidelines [1,2] as follows: rule-in threshold: MRE≥3.3 kPa plus FIB-4≥1.6; and rule-out threshold: MRE<3.3 kPa plus FIB-4<1.6.

Two-gate diagnostic accuracy studies, studies lacking sufficient data to reconstruct 2×2 classification tables, and studies reporting diagnostic accuracy estimates for MEFIB index thresholds other than those prespecified were excluded [11].

Search strategy and study selection

We searched Medline, Cochrane library and Web of Science from inception to February 25th, 2025, without restrictions. We structured our search strategy using free text words and controlled vocabulary (Supplementary Tables 2-4). We used the Polyglot Search Translator to convert search strings across databases [12]. We did not search conference proceedings from relevant scientific meetings.

Search results were imported into reference manager software and duplicates were removed. The remaining records were then imported into the Covidence web application. Pairs of reviewers, working independently, assessed record eligibility, initially at title and abstract level and then in full text. Disagreements were resolved either through discussion between the original reviewers, or by a senior reviewer.

Data extraction and quality assessment

Two reviewers working independently extracted data from eligible studies using predesigned and pilot-tested forms. Data extraction items included study characteristics, participant characteristics and diagnostic accuracy results in terms of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). If raw data for TP, FP, TN, FN were not available in the original studies, we computed them from the sample size, prevalence and other diagnostic accuracy measures using RevMan's calculator. To identify overlapping cohorts among included studies, we took into consideration recruitment periods, participating centers and authors. In case of overlapping cohorts across publications, we prioritized results from the cohorts with the largest sample size, provided they reported sufficient information for 2×2 classification tables.

Two reviewers working independently assessed the risk of bias and applicability of the included studies using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool [13]. Details on risk of bias and applicability judgements are presented in the Supplementary material. Disagreements during the data extraction and quality assessment process were resolved through discussion, or by a senior reviewer.

[a]Clinical Research and Evidence-Based Medicine Unit, Second Medical Department, Aristotle University of Thessaloniki, Greece (Konstantinos Malandris, Aris Liakos, Thomas Karagiannis, Eleni Theocharidou, Apostolos Tsapas); [b]Second Propaedeutic Medical Department, Aristotle University of Thessaloniki, Greece (Anastasia Katsoula); [c]Union Memorial Hospital, Georgetown University, Baltimore, Maryland, USA (Tarek Nayfeh); [d]School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki, Greece (Kalliopi Tsapa, Dimitra Tsapa); [e]First Medical Department, Aristotle University of Thessaloniki, Greece (Georgios Kalopitas, Georgios Germanidis); [f]Fourth Medical Department, Aristotle University of Thessaloniki, Greece (Emmanouil Sinakos); [g]Harris Manchester College, University of Oxford, UK (Apostolos Tsapas)

Data synthesis and statistical analysis

The primary outcomes of interest were the accuracy of the MEFIB index for ruling in at-risk MASH (NAS≥4 and fibrosis stage ≥F2), and significant fibrosis (F≥F2). Secondary outcomes of interest were the accuracy of the MEFIB index for ruling out at-risk MASH, and significant fibrosis.

For all outcomes we reconstructed 2×2 classification tables from eligible studies. Using respective data, we recalculated sensitivity and specificity estimates, with their 95% confidence intervals (CIs), and created coupled forest plots to visually present these estimates. In view of the homogeneity of thresholds for the index test among primary studies, we calculated pooled specificity, sensitivity, positive likelihood ratio (LRp) and negative likelihood ratio (LRn), using the bivariate random-effects model [14,15]. We graphically present individual and pooled study estimates in receiver operating characteristic (ROC) space alongside 95% confidence and prediction regions. We assessed heterogeneity through visual inspection of forest plots and the size of prediction regions [14]. Given the limited number of included studies, we did not investigate for potential sources of heterogeneity through meta-regression analysis [16]. We assessed for the presence of small-study effect bias by means of Deeks' funnel plots, with P<0.10 for the slope coefficient indicating significant asymmetry [17]. We used Cook's distance approach and standardized residuals to identify potentially influential studies (Supplementary material) [18]. We conducted prespecified sensitivity analyses, excluding influential studies identified using Cook's distance approach, studies with unclear or high applicability concerns, and studies of retrospective design, because of potential bias related to the disease spectrum and the overestimation of diagnostic accuracy estimates [19]. All these analyses were conducted solely for the primary outcomes. To assess the clinical utility of the MEFIB index for ruling in at-risk MASH and significant fibrosis we used Fagan nomograms, assuming various Pretest probabilities reflecting both high and low prevalence settings. In addition, using the pooled estimates of sensitivity and specificity, we calculated PPVs and negative predictive values (NPVs) for all outcomes for the same prevalence scenarios. We performed all analyses using STATA statistical software v.11.2 and MetaDTA [20,21].

Results

After removing duplicates, we screened 682 records at title and abstract level, from which 30 full-text articles were assessed for eligibility. Eventually, 7 studies with 3356 participants were included in the systematic review and meta-analysis (Fig. 1) [8,22-27].

Study and participant characteristics

Table 1 presents the main characteristics of the included studies and participants. Most studies were multicenter, following a prospective design, recruiting mainly participants from tertiary healthcare facilities. One study included participants from a low prevalence setting (those referred for routine colorectal cancer screening) [24]. Two studies were identified solely as conference abstracts [24,27]. The study by Loomba *et al* provided the largest amount of data, comprising nearly 2000 participants who were screened for enrollment in the MAESTRO-MASH clinical trial [27]. The mean age of participants ranged from 39.0 to 65.0 years. Among the 3356 participants, almost half (46%) were males and 55.8% (1,872 participants) had type 2 diabetes. The average mean body mass index (BMI) was 30.3 kg/m^2, with a trend towards lower values for Asian cohorts (27.8 kg/m^2). The mean aspartate transaminase (AST) and alanine transaminase (ALT) values ranged from 36.6-56.6 IU/L and from 50.6-84.0 IU/L, respectively. The average mean FIB-4 index was 1.75, ranging from 0.98-2.80. Similarly, the average mean MRE value was 3.6 kPa, ranging from 2.7-5.1 kPa. Among studies with available data, the prevalence of at-risk MASH was 31.3% (393 of 1255 participants), while the prevalence of significant fibrosis was 60.1% (1,916 of 3186 participants).

Risk of bias assessment and applicability

Three studies were at unclear or high risk for bias, because of concerns related to patient selection [23,24,26]. One study raised applicability concerns due to the low prevalence setting from which participants were recruited (during referral for colon cancer screening) [24]. A detailed presentation of risk of bias and applicability assessment is presented in Supplementary Table 5.

Accuracy of MEFIB index for ruling in/out at-risk MASH

Five studies with 1255 participants contributed data to this analysis [8,22-25]. The study by Kim *et al* included 2 different cohorts (USCD cohort and Yokohama cohort), which were handled separately to facilitate analysis [8]. Fig. 2 presents individual study estimates for ruling in at-risk MASH. Sensitivity and specificity estimates across studies ranged from 0.05-0.64 and from 0.63-1.00, respectively. MEFIB index (MRE≥3.3 kPa plus FIB-4≥1.6) yielded a pooled sensitivity of 0.34 (95%CI 0.18-0.55), specificity 0.94 (0.74-0.99), LRp 5.3 (95%CI 1.8-15.7) and LRn 0.71 (95%CI 0.57-0.88).

For ruling out at-risk MASH, individual study estimates for sensitivity and specificity ranged from 0.45-0.93 and from 0.43-0.90, respectively (Supplementary Fig. 1). The MEFIB index (MRE<3.3 kPa plus FIB-4<1.6) yielded a pooled sensitivity of 0.77 (95%CI 0.62-0.88), specificity 0.66 (95%CI 0.49-0.80), LRp 2.3 (95%CI 1.6-3.2), and LRn 0.34 (95%CI 0.23-0.52).

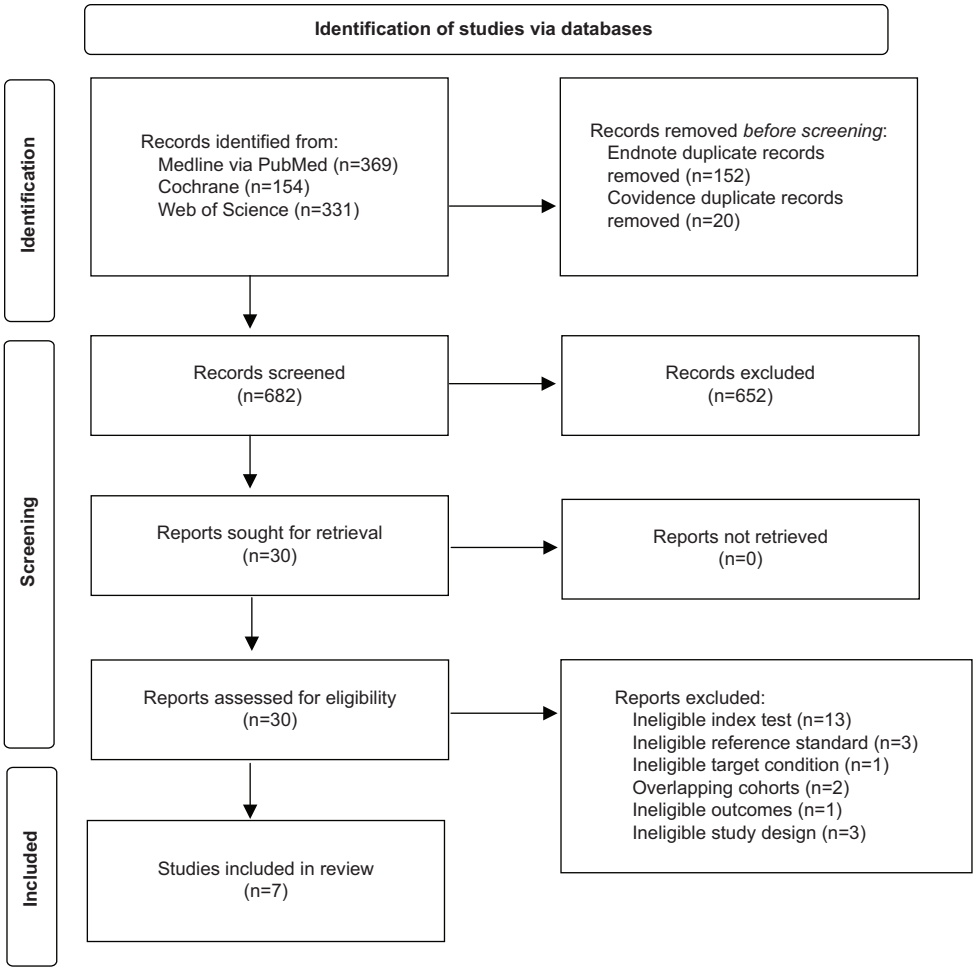


Figure 1 Flow diagram of the study selection process

Accuracy of MEFIB index for ruling in/out significant fibrosis

Fig. 3 presents individual study estimates for ruling in significant fibrosis. Sensitivity and specificity estimates across studies ranged from 0.33-0.88 and from 0.68-0.98, respectively. Based on aggregated data from 4 studies with 2909 participants [8,22,26,27], the MEFIB index (MRE≥3.3 kPa plus FIB-4≥1.6) yielded a pooled sensitivity of 0.56 (95%CI 0.34-0.76), specificity 0.93 (95%CI 0.85-0.97), LRp 8.2 (95%CI 4.5-14.9), and LRn 0.47 (95%CI 0.30-0.75) for ruling in significant fibrosis.

Two studies (808 participants) provided diagnostic accuracy estimates of MEFIB index for ruling out significant fibrosis [8,22]. Individual estimates for sensitivity and specificity ranged from 0.79-0.94 and from 0.66-0.78, respectively (Supplementary Fig. 2). MEFIB index (MRE<3.3 kPa plus FIB-4<1.6) yielded a pooled sensitivity of 0.88 (95%CI 0.79-0.94), specificity 0.73 (95%CI 0.67-0.79), LRp 3.3 (95%CI 2.5-4.3), and LRn 0.16 (95%CI 0.08-0.31) for ruling out significant fibrosis.

Additional analysis

Visual inspection of the forest plots and the size of the prediction regions indicated substantial heterogeneity across all outcomes (Supplementary Fig. 3). To explore for potential sources of heterogeneity for the primary outcomes, we conducted several sensitivity analyses, with results presented in Supplementary Table 6. Specifically, we assessed the impact of excluding studies that: (i) exclusively recruited participants with type 2 diabetes (T2D); (ii) raised applicability concerns; (iii) were conducted retrospectively; and (iv) were deemed influential based on Cook's distance approach and standardized residuals. Across all sensitivity analyses, results remained consistent with our main findings, with specificity estimates exceeding 90% for both primary outcomes. Notably, among the 3356 participants included in our analyses, 2166 were from 2 studies reported as conference abstracts [24,27]. A sensitivity analysis excluding these studies yielded results consistent with the main analyses, with specificity estimates of 0.93 for both primary outcomes. Only 1 study recruited patients with T2D exclusively [22].

Table 1 Baseline characteristics of included studies

Author, Year [ref.]	Country	Centers, design	Participants, N	Males, N (%)	Mean age±SD	Diabetes, N (%)	Mean BMI, kg/m²±SD	Mean ALT, IU/L±SD	Mean AST, IU/L±SD	Mean FIB-4±SD	Mean MRE, kPa±SD	At-risk MASH, N (%)	Significant fibrosis, N (%)	MEFIB index gray zone, N (%)
Castera, 2024 [22]	France	Multicenter, Prospective	245	159 (64.9)	58.6±9.6	245 (100)	31.3±5.2	50.6±25.3	36.6±14.1	1.30±0.6	3.2±0.8	95 (38.8)	128 (52.2)	90 (36.7)
Qi, 2024 [23]	China	Single center, Prospective	108	60 (55.6)	39.0±12.7	26 (24.1)	28.6±4.2	84.0±51.1	56.6±35.3	0.98±0.6	2.7±0.4	28 (25.9)	34 (31.5)	27 (25.0)
Noureddin, 2023 [24]	USA	Multicenter, Prospective	170	101 (59.4)	55.9±6.0	NR	33.1±4.9	NR	NR	NR	NR	20 (11.8)	NR	23 (14.0)
Imajo, 2023[a] [25]	Japan	Multicenter, Retrospective	169	23 (13.6)	61.8±15.5	61 (36.1)	27.7±4.8	64.3±43.4	51.0±28.1	NR	NR	73 (43.2)	103 (60.9)	54 (32.0)
Kim, 2022 [8]	USA	Multicenter, Prospective	249	106 (42.6)	52.6±13.1	100 (40.2)	31.6±4.4	60.5±42.9	42.9±27.3	1.52±1.2	3.0±1.3	53 (21.3)	80 (32.1)	139 (24.7)
	Japan		314	170 (54.1)	59.7±13.2	189 (60.2)	27.9±4.1	56.4±39.9	46.8±25.5	2.50±2.1	4.2±1.8	124 (39.5)	208 (66.2)	
Inada, 2022 [26]	Japan	Single center, Retrospective	105	47 (44.8)	65.0±10.5	53 (50.5)	27.3±3.6	59.3±34.5	55.0±25.5	2.80±1.3	5.1±2.3	NR	80 (76.2)	NR
Loomba, 2022 [27]	International	Multicenter, Prospective	1,996	878 (44.0)	56.8±11.0	1198 (60.0)	35.6±6.8	54.6±33.9	40.1±23.3	1.41±0.7	3.5±1.0	NR	1283 (64.3)	NR

[a]The study by Imajo et al included 3 different cohorts (Japanese primary, Japanese validation, UCSD) providing data on the diagnostic accuracy of MRE combined with the fibrosis-4 index (MEFIB index). We used only the data from the Japanese validation cohort, as there were some concerns about population overlapping between the remaining 2 cohorts and the study by Kim *et al*

At-risk MASH was defined as MASH with NAS ≥4 and fibrosis stage ≥F2. Significant fibrosis was defined as F≥F2

BMI, body mass index; ALT, alanine transaminase; AST, aspartate transaminase; FIB-4, fibrosis index 4; MASH, metabolic dysfunction-associated steatohepatitis; MRE, magnetic resonance elastography; N, number; NR, not reported; SD, standard deviation

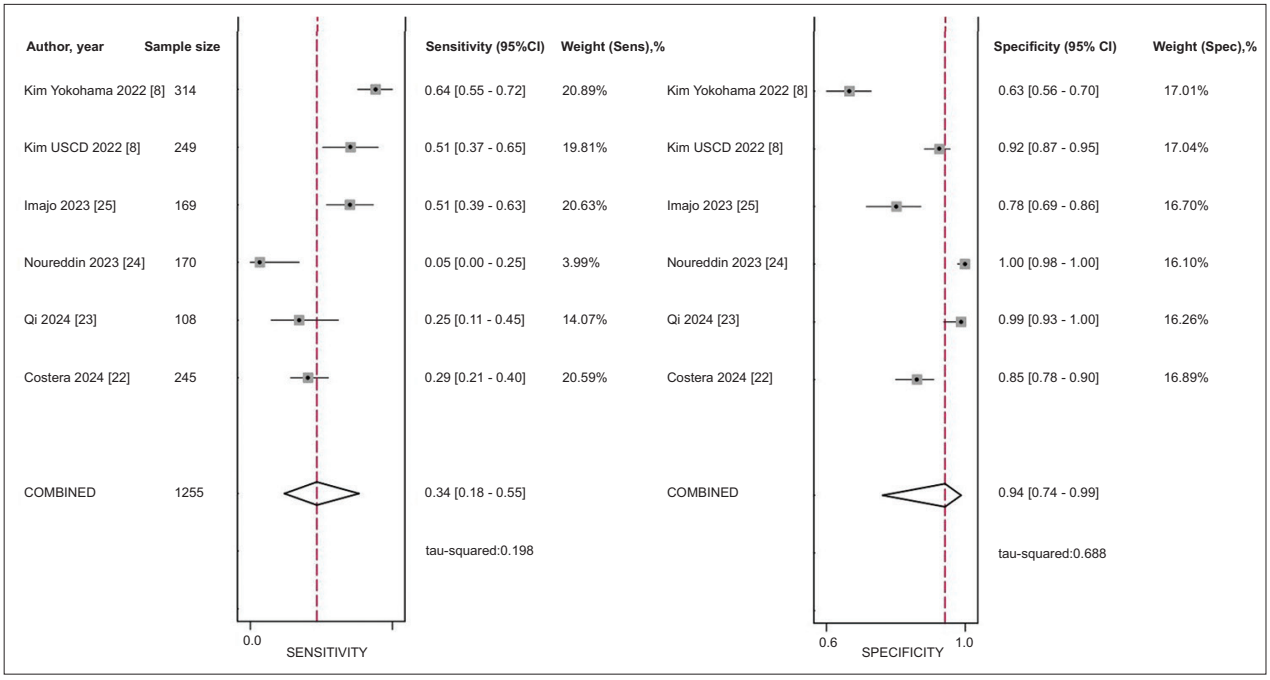


Figure 2 Coupled forest plot of sensitivity and specificity of MEFIB index for ruling in at-risk MASH
MEFIB index, magnetic resonance elastography combined with the fibrosis-4 index; MASH, metabolic dysfunction-associated steatohepatitis; CI, confidence interval

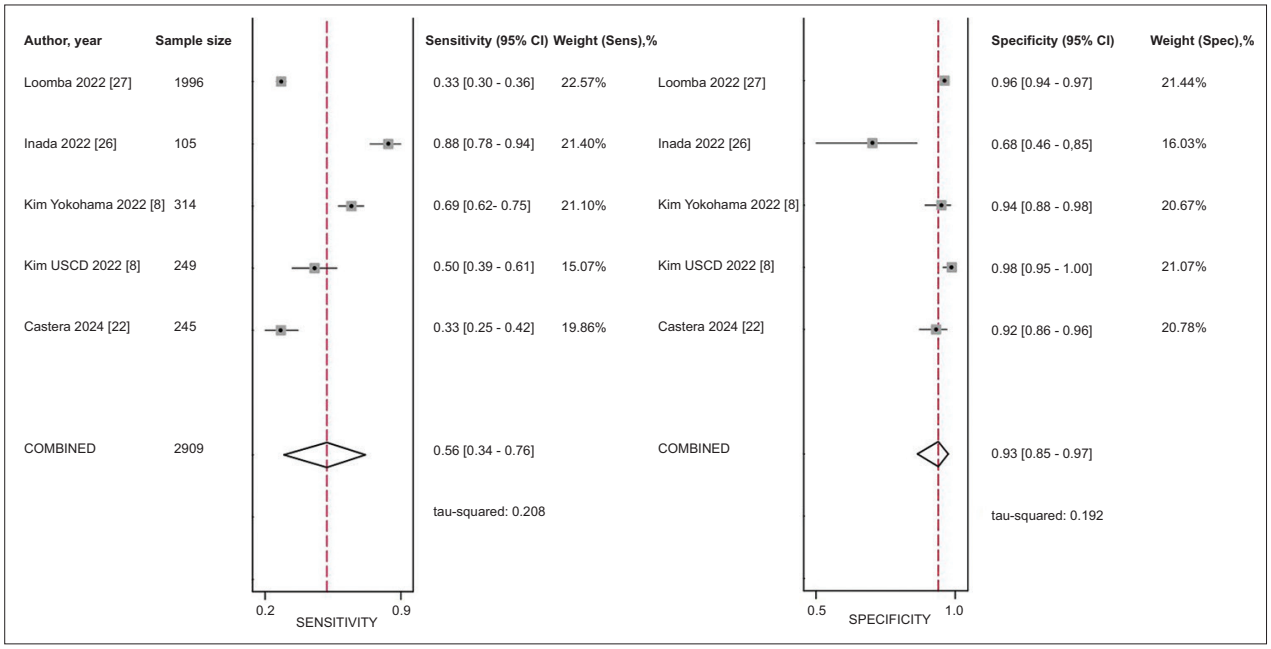


Figure 3 Coupled forest plot of sensitivity and specificity of the MEFIB index for ruling in significant fibrosis
MEFIB index, magnetic resonance elastography combined with the fibrosis-4 index; CI, confidence interval

This study reported specificity estimates of 0.85 and 0.92 for ruling in at risk-MASH and significant fibrosis respectively. In *post hoc* analyses by cohort region, Asian cohorts yielded pooled specificity estimates of 0.87 and 0.92 for at-risk MASH and significant fibrosis, respectively. The respective estimates from Western cohorts were similar at 0.97 and 0.93. In sensitivity analyses including only studies at low risk of bias for all QUADAS domains, pooled specificity estimates for ruling in at-risk MASH and significant fibrosis were 0.82 and 0.96, respectively. The study by Noureddin *et al* was influential for ruling in at-risk MASH, yielding the highest specificity estimate (Supplementary Fig. 4, 5). This study was identified solely as a conference abstract, limiting detailed assessment of baseline characteristics. However, it is worth

mentioning that this study had the lowest prevalence of at-risk MASH among the included studies (11.8%).

Based on Deeks' funnel plots (Supplementary Fig. 6, 7) there was evidence of small study effect bias for ruling in at risk MASH (P=0.02). However, the number of included studies was limited, thus limiting the reliability of the respective analysis.

Clinical utility

Assuming a prevalence of 10-50%, the probability of having at-risk MASH following a positive test was 37-84%, respectively (Supplementary Fig. 8). For a prevalence of 60-80%, the probability for at-risk MASH increased, ranging from 89-96% respectively. For ruling in significant fibrosis, and for a prevalence setting ranging from 10-50%, the post-test probability after a positive test result ranged from 48-89%, respectively (Supplementary Fig. 9). For higher prevalence settings (60-80%), respective post-test probabilities for having significant fibrosis ranged from 92-97%. Table 2 presents PPVs and NPVs of the MEFIB index for all outcomes and for the same prevalence scenarios.

Discussion

In this systematic review and meta-analysis, we evaluated the accuracy of the MEFIB index for identifying at-risk MASH and significant fibrosis in adults with MASLD, using biopsy as the reference standard. We limited our analysis to the thresholds recommended by relevant societies: MRE≥3.3 kPa plus FIB-4≥1.6 to rule in the target conditions, and MRE<3.3 kPa plus FIB-4<1.6 to rule them out. Given that the MEFIB index was developed to address the low PPVs of existing noninvasive tests, and to facilitate participant selection for clinical trials, our analysis primarily focused on assessing its accuracy in diagnosing at-risk MASH and significant fibrosis.

Table 2 Positive and negative predictive values for all outcomes across different prevalence scenarios

Prevalence	At-risk MASH		Significant fibrosis	
	PPVs for ruling in	NPVs for ruling out	PPVs for ruling in	NPVs for ruling out
10%	39%	96%	47%	98%
20%	59%	92%	67%	96%
30%	71%	87%	77%	93%
40%	79%	81%	84%	90%
50%	85%	74%	89%	86%
60%	89%	66%	92%	80%
70%	93%	55%	95%	72%
80%	96%	42%	97%	60%

PPV, positive predictive value; NPV, negative predictive value. PPV and NPV values in this table derive from pooled sensitivity and specificity estimates

Based on our findings, the MEFIB index demonstrated robust performance in identifying both target conditions. For ruling in at-risk MASH, MEFIB index achieved a pooled specificity of 0.94 and an LRp of 5.3. Similarly, for ruling in significant fibrosis, the index yielded a summary specificity of 0.93 and an LRp of 8.2. In a prevalence setting of 60%, the MEFIB index resulted in a PPV exceeding 90% for significant fibrosis and 89% for at-risk MASH. For ruling out the target conditions, the MEFIB index yielded pooled sensitivity estimates of 0.77 for at-risk MASH and 0.88 for significant fibrosis.

Our systematic review and meta-analysis provides a timely placed synthesis of evidence concerning the diagnostic performance of the MEFIB index. Using robust methodology, in line with Cochrane recommendations, we searched several databases and included 7 studies with more than 3000 participants. Our clinically focused results employed the dual cutoff approach, using the most widely used MEFIB index thresholds for ruling in or ruling out at-risk MASH and significant fibrosis. By focusing on specific MEFIB index positivity thresholds, we were able to provide summary estimates of sensitivity and specificity—metrics that offer greater clinical utility than the less informative area under the ROC curve (AUROC). Furthermore, for at-risk MASH, we employed the definition most commonly used for patient selection in MASH clinical trials. This choice was made to maximize the external validity and translatability of our findings to clinical trial settings and real-world practice.

Certain limitations must be acknowledged. Visual inspection of forest plots and the size of prediction regions indicated high heterogeneity for all outcomes of interest. Given the limited number of studies included in our meta-analysis (7 studies), we were unable to assess for potential sources of heterogeneity through meta-regression analysis [16]. Nevertheless, several exploratory sensitivity analyses were conducted, with results consistent with our main findings. Sparse reporting of relevant data prevented us from performing subgroup analyses based on specific factors previously suggested to influence the diagnostic performance of newly developed noninvasive biomarkers, including T2D, BMI, and age (≥65 years). Additionally, most included studies were at unclear or high risk for bias, primarily due to patient selection concerns. This was mainly attributable to the retrospective design of the studies and the possibility of convenience sampling, or suboptimal reporting of enrolment procedures.

To our knowledge, this is the first meta-analysis evaluating the diagnostic performance of the MEFIB index for identifying or excluding at-risk MASH and significant fibrosis. For ruling in significant fibrosis, our findings (specificity: 0.93) closely align with the results reported by Kim *et al* [8]. In their study, Kim *et al* combined 2 geographically distinct cohorts—a testing cohort from the USA (UCSD) (specificity: 0.98) and a validation cohort from Japan (Yokohama) (specificity: 0.94). For ruling in at-risk MASH, our pooled specificity estimates significantly differed from the combined estimates reported by Kim *et al* (0.94 vs. 0.77). Notably, our specificity estimates closely match that reported by the UCSD cohort alone (0.94 vs. 0.91), while the primary discrepancy arises from the Yokohama cohort, which reported a considerably lower specificity of 0.63. Although the Yokohama cohort had a lower mean BMI compared to our study population (27.9 vs. 30.3 kg/m^2), we do not consider this difference in BMI

as the main reason for the observed discrepancy. Emerging evidence from well-conducted individual patient data meta-analyses suggests that BMI does not substantially confound MRE metrics in MASLD [28], thus highlighting the need for further validation of the MEFIB index in other cohorts.

Recently, the MRI-AST (MAST) score, combining MRI-proton density fat fraction (PDFF), MRE, and AST levels, was introduced for diagnosing at-risk MASH [29]. When comparing MEFIB and MAST directly, MEFIB appears superior based on AUROC comparisons; however, the MAST score has the advantage of yielding a lower percentage of unclassified participants (gray zone) [8]. Specifically, the pooled prevalence of gray zone results for MEFIB index in our analysis was 26.5%, whereas the respective reported prevalence for the MAST score is 18.1% [29]. The FAST score is another noninvasive biomarker that was recently developed in order to facilitate patient selection for clinical trials [30]. Published meta-analyses report a FAST score specificity of around 0.90 for ruling in at risk MASH, with a PPV of 87% for a prevalence of 60% [31,32]. Nevertheless, results from comparative diagnostic accuracy studies support the superiority of the MEFIB index over the FAST score in terms of AUROC comparison (0.76 vs. 0.68), with similar gray zone magnitudes (26.1% vs. 30.8%) [8]. On the other hand, the FAST score offers the advantages of lower cost and easier applicability compared to an MRI examination. A structured comparison between FAST, MEFIB and MAST score is presented in Supplementary Table 7.

Early identification of at-risk MASH or significant fibrosis is important for timely initiation of appropriate pharmacotherapy, intensification of comorbidity management and close monitoring for disease progression. With a pooled specificity of 0.94, the MEFIB index accurately classifies nearly 9 of 10 patients with at-risk MASH. Similarly, with a summary specificity of 0.93, MEFIB reliably identifies approximately 9 of 10 patients without significant fibrosis, yielding roughly 1 false positive per 10 patients tested. In addition, a positive MEFIB result indicates that patients are approximately 5 times more likely to have at-risk MASH (LRp 5.3) and nearly 8 times more likely to have significant fibrosis (LRp 8.2) compared to those testing negative. As a result, it seems that MEFIB performs better for diagnosing significant fibrosis compared to at-risk MASH. This might be related to the fact that both MEFIB components mainly target fibrosis rather than other histological features of MASH, such as steatosis, inflammation and ballooning.

It should be noted that a substantial proportion of patients initially classified within the low or indeterminate risk categories based on FIB-4 scores have subsequently been identified as having clinically significant fibrosis [33]. As a result, a low FIB-4 during MEFIB should be followed by further examination and diagnostic evaluation in the presence of clinical uncertainty. Noureddin *et al* provide an example of such a case, where a 50-year-old patient with MASLD had AST 45 U/L, ALT 60 U/L, platelet count 270×10^9/L, MRI-PDFF 15%, MRE 4 kPa, controlled attenuation parameter 345 dB/m, and VCTE 12 kPa [34]. This patient would have a FIB-4 of score of 1.08, while his FAST and MAST scores suggest the presence of at-risk MASH [34]. As a result, MEFIB, MAST and FAST should not be considered as competing candidates, rather as useful tools in the holistic evaluation of a patient with MASLD.

Similarly to other scores utilizing a dual cutoff approach, the MEFIB index suffers the limitation of gray zone results (26% of participants). Assessment of these patients should be done by taking into account proximity to thresholds, patient characteristics, and additional testing by means of other noninvasive scores, before liver biopsy. Notably, a recently published meta-analysis of individual participant data found that a positive MEFIB index had a strong association with liver-related outcomes, hepatocellular carcinoma and death, and a high NPV of 99% for hepatic decompensation at 5 years [35,36].

Limitations in the diagnostic accuracy, availability and cost of current noninvasive tests have led to recommendations advocating for their sequential application. This strategy typically begins with tests that are widely accessible and easy to apply, followed by more specialized ones [37]. Although various combinations of tests may be employed, the underlying principle remains the same: increasing the prevalence of the target condition within the tested population to enhance the PPV of the subsequent test.

In conclusion, the MEFIB index has acceptable accuracy for diagnosing at-risk MASH and significant fibrosis. The proposed thresholds can be used to identify both target conditions in high prevalence settings, and to facilitate patient recruitment in clinical trials.

Summary Box

What is already known:

- Patients with metabolic dysfunction-associated steatohepatitis ([MASH], nonalcoholic fatty liver disease activity score ≥4) and significant fibrosis (≥F2) (at-risk MASH) are at increased risk for disease progression
- Magnetic resonance elastography (MRE) combined with the fibrosis-4 index (MEFIB index) enables the noninvasive diagnosis of at-risk MASH and significant fibrosis
- The MEFIB index was originally developed to address the low positive predictive values (PPVs) of existing noninvasive tests, and to facilitate participant selection for clinical trials

What the new findings are:

- For ruling in at-risk MASH, the MEFIB index achieved a pooled specificity of 0.94 and a positive likelihood ratio (LRp) of 5.3
- For ruling in significant fibrosis, the index yielded a summary specificity of 0.93 and an LRp of 8.2
- In a prevalence setting of 60%, the MEFIB index resulted in a PPV exceeding 90% for significant fibrosis and 89% for at-risk MASH
- The MEFIB index can be used to identify both target conditions in high prevalence settings, and to facilitate patient recruitment in clinical trials

References

1. Rinella ME, Neuschwander-Tetri BA, Siddiqui MS, et al. AASLD Practice Guidance on the clinical assessment and management of nonalcoholic fatty liver disease. *Hepatology* 2023;**77**:1797-1835.
2. European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines on the management of metabolic dysfunction-associated steatotic liver disease (MASLD). *J Hepatol* 2024;**81**:492-542.
3. Younossi ZM, Stepanova M, Ong J, et al. Nonalcoholic steatohepatitis is the most rapidly increasing indication for liver transplantation in the United States. *Clin Gastroenterol Hepatol* 2021;**19**:580-589.
4. Harrison SA, Allen AM, Dubourg J, Noureddin M, Alkhouri N. Challenges and opportunities in NASH drug development. *Nat Med* 2023;**29**:562-573.
5. Zoncapè M, Liguori A, Tsochatzis EA. Non-invasive testing and risk-stratification in patients with MASLD. *Eur J Intern Med* 2024;**122**:11-19.
6. Younossi ZM, Zelber-Sagi S, Lazarus JV, et al. Global consensus recommendations for metabolic dysfunction-associated steatotic liver disease and steatohepatitis. *Gastroenterology* 2025;**11**:S0016-5085(25)00632-8.
7. Brennan PN, Kopka CJ, Agirre-Garrido L, et al. Reviewing MAESTRO-NASH and the implications for hepatology and health systems in implementation/accessibility of Resmetirom. *NPJ Gut Liver* 2025;**2**:3.
8. Kim BK, Tamaki N, Imajo K, et al. Head-to-head comparison between MEFIB, MAST, and FAST for detecting stage 2 fibrosis or higher among patients with NAFLD. *J Hepatol* 2022;**77**:1482-1490.
9. 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.
10. McInnes MDF, Moher D, Thombs BD, et al; and the PRISMA-DTA Group. Preferred Reporting Items for a Systematic review and Meta-Analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA* 2018;**319**:388-396.
11. Bossuyt PM. Chapter 3: Understanding the design of test accuracy studies. In: Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y (editors). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 2.0 (updated July 2023). Cochrane, 2023. Available from: https://training.cochrane.org/handbook-diagnostic-test-accuracy/current [Accessed 23 September 2025].
12. Clark JM, Sanders S, Carter M, et al. Improving the translation of search strategies using the Polyglot Search Translator: a randomized controlled trial. *J Med Libr Assoc* 2020;**108**:195-207.
13. Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;**155**:529-536.
14. Macaskill P, Takwoingi Y, Deeks JJ, Gatsonis C. Chapter 9: Understanding meta-analysis. In: Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y (editors). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 2.0 (updated July 2023). Cochrane, 2023. Available from: https://training.cochrane.org/handbook-diagnostic-test-accuracy/current [Accessed 23 September 2025].
15. Takwoingi Y, Dendukuri N, Schiller I, Rücker G, Jones HE, Partlett C, Macaskill P. Chapter 10: Undertaking meta-analysis. In: Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y (editors). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 2.0 (updated July 2023). Cochrane, 2023. Available from: https://training.cochrane.org/handbook-diagnostic-test-accuracy/current [Accessed 23 September 2025].
16. Deeks JJ, Higgins JPT, Altman DG, McKenzie JE, Veroniki AA (editors). Chapter 10: Chapter 10: Analysing data and undertaking meta-analyses [last updated November 2024]. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.5. Cochrane, 2024. Available from: https://www.training.cochrane.org/handbook [Accessed 23 September 2025].
17. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 2005;**58**:882-893.
18. Skrondal A, Rabe-Hesketh, S. Generalized latent variable modeling: multilevel, longitudinal, and structural equation models (1st ed.). Chapman and Hall/CRC, 2004. doi: 10.1201/9780203489437
19. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006;**174**:469-476.
20. Freeman SC, Kerby CR, Patel A, Cooper NJ, Quinn T, Sutton AJ. Development of an interactive web-based tool to conduct and interrogate meta-analysis of diagnostic test accuracy studies: MetaDTA. *BMC Med Res Methodol* 2019;**19**:81.
21. Patel A, Cooper N, Freeman S, Sutton A. Graphical enhancements to summary receiver operating characteristic plots to facilitate the analysis and reporting of meta-analysis of diagnostic test accuracy data. *Res Synth Methods* 2021;**12**:34-44.
22. Castera L, Garteiser P, Laouenan C, et al; QUID NASH investigators. Prospective head-to-head comparison of non-invasive scores for diagnosis of fibrotic MASH in patients with type 2 diabetes. *J Hepatol* 2024;**81**:195-206.
23. Qi S, Wei X, Zhao J, et al. Performance of MAST, FAST, and MEFIB in predicting metabolic dysfunction-associated steatohepatitis. *J Gastroenterol Hepatol* 2024;**39**:1656-1662.
24. Noureddin M, Alkhouri N, Chaldaureille C, et al. Head-to-head comparison of FAST, MAST, MEFIB, and cT1 in identifying at-risk NASH patients in a low-prevalence population. *Hepatology* 2023;**78**:S806-S810.
25. Imajo K, Saigusa Y, Kobayashi T, et al. M-PAST score is better than MAST score for the diagnosis of active fibrotic nonalcoholic steatohepatitis. *Hepatol Res* 2023;**53**:844-856.
26. Inada K, Tamaki N, Kurosaki M, et al. Validation of magnetic resonance elastography plus fibrosis-4 for significant fibrosis in nonalcoholic fatty liver disease. *J Gastroenterol Hepatol* 2022;**37**:1726-1731.
27. Loomba R, Harrison SA, Taub RA, et al. 102 | Utility of FIB-4, MRE, MRI-PDFF, and FibroScan to identify patients with at-risk F2-F3 NASH based on screening data from a 2000 patient biopsy confirmed cohort of resmetirom phase 3 clinical trial, MAESTRO-NASH. *Hepatology (Baltimore, Md)* 2022;**76** (Suppl 1):S92-S94.
28. Liang JX, Ampuero J, Niu H, et al; LITMUS Consortium Investigators. An individual patient data meta-analysis to determine cut-offs for and confounders of NAFLD-fibrosis staging with magnetic resonance elastography. *J Hepatol* 2023;**79**:592-604.
29. Noureddin M, Truong E, Gornbein JA, et al. MRI-based (MAST) score accurately identifies patients with NASH and significant fibrosis. *J Hepatol* 2022;**76**:781-787.
30. 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**:362-373.
31. Malandris K, Arampidis D, Mainou M, et al. FibroScan-AST score for diagnosing fibrotic MASH: a systematic review and meta-analysis of diagnostic test accuracy studies. *J Gastroenterol Hepatol* 2024;**39**:2582-2591.
32. Ravaioli F, Dajti E, Mantovani A, Newsome PN, Targher G, Colecchia A. Diagnostic accuracy of FibroScan-AST (FAST) score

for the non-invasive identification of patients with fibrotic non-alcoholic steatohepatitis: a systematic review and meta-analysis. *Gut* 2023;**72**:1399-1409.
33. Hagström H, Talbäck M, Andreasson A, Walldius G, Hammar N. Repeated FIB-4 measurements can help identify individuals at risk of severe liver disease. *J Hepatol* 2020;**73**:1023-1029.
34. Noureddin M, Harrison SA, Alkhouri N. MEFIB vs. MAST and FAST: Not a competition but useful tools. *J Hepatol* 2024;**80**:e35-e36.
35. Tincopa MA, Loomba R. Non-invasive diagnosis and monitoring of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis. *Lancet Gastroenterol Hepatol* 2023;**8**:660-670.
36. Ajmera V, Kim BK, Yang K, et al. Liver stiffness on magnetic resonance elastography and the MEFIB index and liver-related outcomes in nonalcoholic fatty liver disease: a systematic review and meta-analysis of individual participants. *Gastroenterology* 2022;**163**:1079-1089.
37. Tsochatzis EA. Screening for liver fibrosis—sequential non-invasive testing works best. *J Hepatol* 2023;**79**:263-265.

Supplementary material

Risk of bias and applicability assessment

Two reviewers working independently assessed the risk of bias and applicability of included studies using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool. We took into consideration the following domains: patient selection, index test, reference standard, and flow and timing.

DOMAIN 1: PATIENT SELECTION

Signaling question 1*: Was a consecutive or random sample of patients enrolled?*

- Answer Yes if: a consecutive or random sample of patients was enrolled
- Answer Unclear if: not enough information to make a decision
- Answer No if: neither a consecutive nor a random sample of patients was enrolled, and in case of convenience samples (i.e, studies that searched records of patients who had undergone both MEFIB and liver biopsy), retrospective studies

Signaling question 2*: Was a case-control design avoided?*

- Answer Yes if: the study recruited a single group of patients
- Answer No if: the study recruited multiple groups with different inclusion criteria
- Answer Unclear if: not enough information to make a decision

Signaling question 3: *Did the study avoid inappropriate exclusions?*

- Answer Yes if: all patients at risk were included
- Answer No if: specific patient subgroups were excluded for the study
- Answer Unclear: not enough information to make a decision

DOMAIN 2: INDEX TEST

Signaling question 1*: Were the index test results interpreted without knowledge of the results of the reference standard?*

- Answer Yes/No if: MEFIB score components were obtained without/with knowledge to biopsy results
- Answer Unclear if: not enough information to make a decision

Signaling question 2*: If a threshold was used, was it prespecified?*

- Answer Yes if: MEFIB cutoffs were specified
- Answer Unclear: not enough information to make a decision
- Answer No: If the positivity threshold was based on data collected during the study (i.e., Youden index)

DOMAIN 3: REFERENCE STANDARD

Signaling question 1: Is the reference standard likely to correctly classify the target condition?

- Answer Yes if: liver biopsy was used as the reference standard
- Answer No: in any other case

Signaling question 2*: Were the reference standard results interpreted without knowledge of the results of the index test?*

- Answer Yes/No if: biopsy was performed without/with knowledge to index test results
- Answer Unclear if: not enough information to make a decision

DOMAIN 4: FLOW AND TIMING

Signaling question 1*: Was there an appropriate interval between index test and reference standard?*

- Answer Yes if: time interval between liver biopsy and MEFIB ≤ 3months
- Answer No if: time interval between liver biopsy and MEFIB > 6months
- Answer Unclear if: not enough information to make a decision

Signaling question 2*: Did all participants receive a reference standard?*

- Answer Yes if: If all participants received a reference standard
- Answer No if: If not all participants received a reference standard
- Answer Unclear if: not enough information to make a decision

Signaling question 3*: Did all patients receive the same reference standard?*

- Answer Yes if: If all participants received the same reference standard
- Answer No if: If some participants received a different reference standard
- Answer Unclear if: not enough information to make a decision

Signaling question 4*: Were all patients included in the analysis?*

- Answer Yes if: the number of patients enrolled (i.e the number of patients in the baseline table) is same with the number of patients in the 2x2 tables.
- Answer No: if the number of enrolled patients is different from the number of patients included in the 2x2 tables
- Answer Unclear if: not enough information to make a decision

APPLICABILITY CONCERNS

Are there concerns that the included patients do not match the review question?

- Answer No if: all included patients had MASLD
- Answer Yes: in case of other liver diseases
- Answer Unclear if: not enough information to make a decision

Are there concerns that the index test, its conduct, or interpretation differ from the review question?

- Answer No if: MRE was conducted based on a prespecified protocol and FIB-4 was calculated based on the standard formulae
- Answer Unclear if: not enough information to make a decision
- Answer Yes: In any other case

Are there concerns that the target condition as defined by the reference standard does not match the review question?

- Answer No if biopsy was performed in order to detect target condition
- Answer Unclear if: not enough information to make a decision
- Answer Yes: In any other case

Cooks distance and standardized residuals

Cook's distance plots and standardized residuals are best interpreted together. Cook's distance plots help identify potentially influential studies, defined as those exceeding a specified threshold (indicated by a red line). This threshold is calculated by multiplying the number of estimated parameters (n=5; sensitivity, specificity, variance of sensitivity, variance of specificity, and correlation between variances) by 4 and then dividing this product by the total number of studies. For standardized residuals, thresholds of -2 and +2 were applied to assess whether studies had a notable negative or positive influence on sensitivity and/or specificity, respectively.

Supplementary Table 1 PRISMA checklist

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
		TITLE / ABSTRACT	
Title	1	Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies	1
Abstract	2	Abstract: See PRISMA-DTA for abstracts	3
		INTRODUCTION	
Rationale	3	Describe the rationale for the review in the context of what is already known.	4
Clinical role of index test	D1	State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design)	4
Objectives	4	Provide an explicit statement of question(s) being addressed in terms of participants, index test(s), and target condition(s)	4
		METHODS	
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number	5
Eligibility criteria	6	Specify study characteristics (participants, setting, index test(s), reference standard(s), target condition(s), and study design) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale	5
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched	5
Search	8	Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated	Supplementary Tables 2-4
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis)	5,6

(*Contd...*)

Supplementary Table 1 (*Continued*)

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators	6
Definitions for data extraction	11	Provide definitions used in data extraction and classifications of target condition(s), index test(s), reference standard(s) and other characteristics (e.g., study design, clinical setting)	5,6
Risk of bias and applicability	12	Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question	6, Supplementary 1.1
Diagnostic accuracy measures	13	State the principal diagnostic accuracy measure(s) reported (e.g., sensitivity, specificity) and state the unit of assessment (e.g., per-patient, per-lesion)	6,7
Synthesis of results	14	Describe methods of handling data, combining results of studies and describing variability between studies. This could include, but is not limited to: a) handling of multiple definitions of target condition. b) handling of multiple thresholds of test positivity, c) handling multiple index test readers, d) handling of indeterminate test results, e) grouping and comparing tests, f) handling of different reference standards	6,7
Meta-analysis	D2	Report the statistical methods used for meta-analyses, if performed	7
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified	7
		RESULTS	
Study selection	17	Provide numbers of studies screened, assessed for eligibility, included in the review (and included in meta-analysis, if applicable) with reasons for exclusions at each stage, ideally with a flow diagram	7, Figure 1
Study characteristics	18	For each included study provide citations and present key characteristics including: a) participant characteristics (presentation, prior testing), b) clinical setting, c) study design, d) target condition definition, e) index test, f) reference standard, g) sample size, h) funding sources	Table 1
Risk of bias and applicability	19	Present evaluation of risk of bias and concerns regarding applicability for each study	8, Supplementary Table 5
Results of individual studies	20	For each analysis in each study (e.g., unique combination of index test, reference standard, and positivity threshold) report 2x2 data (TP, FP, FN, TN) with estimates of diagnostic accuracy and confidence intervals, ideally with a forest or receiver operator characteristic (ROC) plot	Figures 2-3
Synthesis of results	21	Describe test accuracy, including variability; if meta-analysis was done, include results and confidence intervals.	9
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression; analysis of index test: failure rates, proportion of inconclusive results, adverse events)	9,10
		DISCUSSION	
Summary of evidence	24	Summarize the main findings including the strength of evidence	11
Limitations	25	Discuss limitations from included studies (e.g., risk of bias and concerns regarding applicability) and from the review process (e.g., incomplete retrieval of identified research)	11,12
Conclusions	26	Provide a general interpretation of the results in the context of other evidence. Discuss implications for future research and clinical practice (e.g., the intended use and clinical role of the index test)	12-14
		FUNDING	
Funding	27	For the systematic review, describe the sources of funding and other support and the role of the funders	2

Supplementary Table 2 Medline via PubMed

Search term	Result
1. "MRE"[All Fields] AND "FIB-4"[All Fields]	64
2. MEFIB[tiab]	16
3. MRE [tiab]	3,426
4. Magnetic Resonance Elastography [tiab]	1,534
5. 1 OR 2 OR 3 OR 4	3,903
6. "MASH"[All Fields]	3,855
7. "MASLD"[All Fields]	2,753
8. "NASH"[All Fields]	27,913
9. "NAFLD"[All Fields]	29,773
10. "metabolic dysfunction-associated steatohepatitis"[All Fields]	910
11. "metabolic dysfunction-associated steatotic liver disease"[All Fields]	2,593
12. "non-alcoholic steatohepatitis"[All Fields]	7,156
13. "non-alcoholic fatty liver disease"[All Fields]	36,785
14. "fatty liver"[All Fields]	70,857
15. "steatotic liver disease"[All Fields]	3,011
16. 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15	93,013
17. 5 AND 16	369

Supplementary Table 3 Web of science

Search term	Result
1. ALL ("MEFIB")	30
2. ALL ("MRE")	8,389
3. ALL ("MRE" AND "FIB-4")	63
4. 1-3/OR	8,404
5. ALL ("steatotic liver disease")	3,351
6. ALL ("fatty liver")	72,125
7. ALL ("non-alcoholic fatty liver disease")	22,572
8. ALL ("non-alcoholic steatohepatitis")	8,288
9. ALL ("metabolic dysfunction-associated steatotic liver disease")	2,768
10. ALL ("metabolic dysfunction-associated steatohepatitis")	966
11. ALL ("NAFLD")	33,868
12. ALL ("NASH")	84,447
13. ALL ("MASLD")	3,024
14. ALL ("MASH")	9,280
15. 5-14/OR	163,016
16. 4 AND 15	331

Supplementary Table 4 Cochrane library

Search term	Result
1. MeSH descriptor: [Elasticity Imaging Techniques] explode all trees	266
2. MRE	240
3. MEFIB	1
4. MRE AND FIB-4	18
5. #1 OR #2 OR #3 OR #4	494
6. "steatotic liver disease"	119
7. "fatty liver"	5,988
8. MeSH descriptor: [Non-alcoholic Fatty Liver Disease] explode all trees	1,924
9. "NAFLD"	3,044
10. "NASH"	3,046
11. "MASLD"	129
12. "MASH"	217
13. "metabolic dysfunction-associated steatotic liver disease"	108
14. "metabolic dysfunction-associated steatohepatitis"	83
15. #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14	7,883
16. #5 AND #15	154

Supplementary Table 5 Risk of bias and applicability concerns of included studies

	Risk of bias assessment				Applicability assessment		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Castera, 2024	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Qi, 2024	Unclear	Low risk	Unclear	Low risk	Low risk	Low risk	Low risk
Noureddin, 2023	Unclear	Low risk	Low risk	Unclear	High risk	Low risk	Low risk
Imajo, 2023	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Kim, 2022	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Inada, 2022	High risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Loomba, 2022	Low risk	Low risk	High risk	Unclear	Low risk	Low risk	Low risk

Supplementary Table 6 Results from sensitivity analyses

Target Condition	N. of studies	N. of participants	Sensitivity (95%CI)	Specificity (95%CI)	LRp (95%CI)	LRn (95%CI)
Rule in at-risk MASH						
Main analysis	5	1255	0.34 (0.18 to 0.55)	0.94 (0.74 to 0.99)	5.3 (1.8 to 15.7)	0.71 (0.57 to 0.88)
Excluding studies recruiting solely patients with T2D	4	1010	0.36 (0.17 to 0.60)	0.96 (0.70 to 0.99)	9.0 (1.7 to 48.9)	0.67 (0.51 to 0.88)
Excluding studies with applicability concerns	4	1085	0.29 (0.11 to 0.56)	0.93 (0.69 to 0.99)	4.2 (1.4 to 12.7)	0.77 (0.60 to 0.97)
Excluding studies of retrospective design	4	1086	0.35 (0.16 to 0.61)	0.91 (0.66 to 0.98)	4.0 (1.5 to 10.8)	0.71 (0.55 to 0.92)
Excluding influential studies	4	1085	0.29 (0.11 to 0.56)	0.93 (0.69 to 0.99)	4.2 (1.4 to 12.7)	0.77 (0.60 to 0.97)
Only studies at low risk of bias for all QUADAS domains	3	977	0.48 (0.36 to 0.61)	0.82 (0.69 to 0.90)	2.6 (1.6 to 4.1)	0.64 (0.52 to 0.78)
Only Asian cohorts	3	591	0.46 (0.26 to 0.67)	0.87 (0.54 to 0.97)	3.6 (1.2 to 11.0)	0.13 (0.03 to 0.46)
Only Western cohorts	3	664	0.23 (0.06 to 0.56)	0.97 (0.68 to 0.99)	8.3 (1.1 to 63.2)	0.79 (0.59 to 1.05)
Excluding studies published as conference abstracts	4	1085	0.29 (0.11 to 0.56)	0.93 (0.69 to 0.99)	4.2 (1.4 to 12.7)	0.77 (0.60 to 0.97)
Rule in significant fibrosis						
Main analysis	4	2909	0.56 (0.34 to 0.76)	0.93 (0.85 to 0.97)	8.2 (4.5 to 14.9)	0.47 (0.30 to 0.75)
Excluding studies recruiting solely patients with T2D	3	2664	0.62 (0.37 to 0.82)	0.93 (0.82 to 0.98)	9.0 (4.3 to 18.9)	0.41 (0.23 to 0.72)
Excluding studies with applicability concerns	NA	NA	NA	NA	NA	NA
Excluding studies of retrospective design	3	2804	0.58 (0.31 to 0.81)	0.91 (0.82 to 0.96)	6.7 (4.3 to 10.4)	0.46 (0.25 to 0.83)
Excluding influential studies	NA	NA	NA	NA	NA	NA
Only studies at low risk of bias for all QUADAS domains	2	808	0.51 (0.33 to 0.68)	0.96 (0.92 to 0.98)	11.5 (5.0 to 26.0)	0.51 (0.35 to 0.75)
Only Asian cohorts	2	419	0.72 (0.40 to 0.91)	0.92 (0.52 to 0.99)	8.7 (1.4 to 52.1)	0.30 (0.13 to 0.69)
Only Western cohorts	2	494	0.51 (0.27 to 0.75)	0.93 (0.89 to 0.96)	7.7 (3.2 to 18.5)	0.52 (0.30 to 0.91)
Excluding studies published as conference abstracts	3	913	0.62 (0.38 to 0.81)	0.93 (0.80 to 0.97)	8.4 (3.6 to 19.2)	0.41 (0.23 to 0.71)

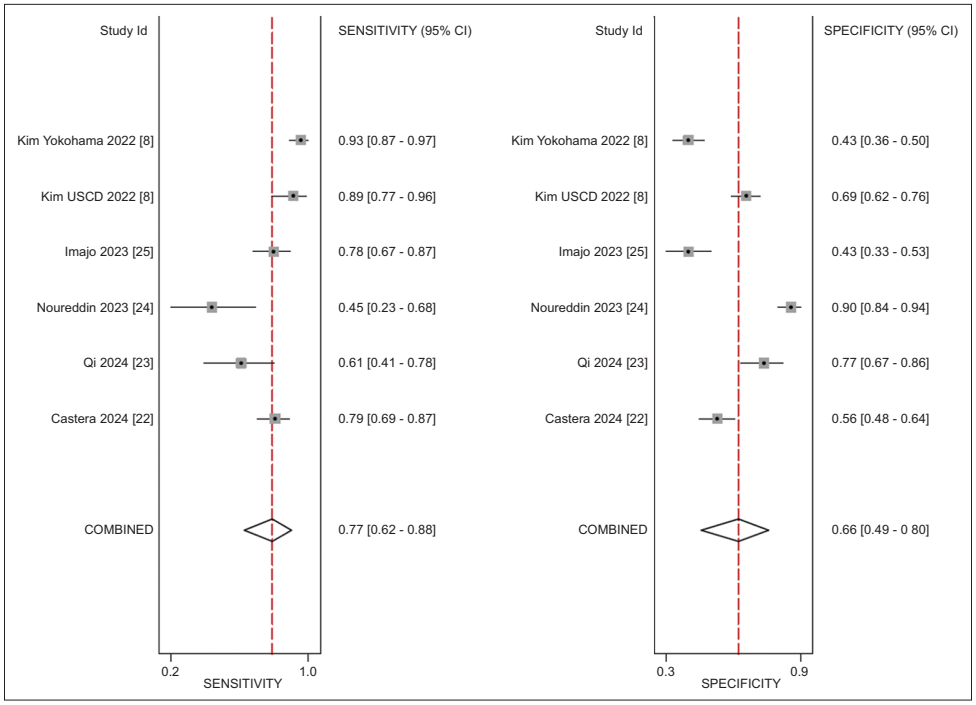
N, Number; NA, Not applicable; CI, Confidence interval; LRp, positive likelihood ratio; LRn, negative likelihood ratio; T2D, Type 2 diabetes

Supplementary Table 7 Comparison between MEFIB, FAST and MAST scores for at-risk MASH

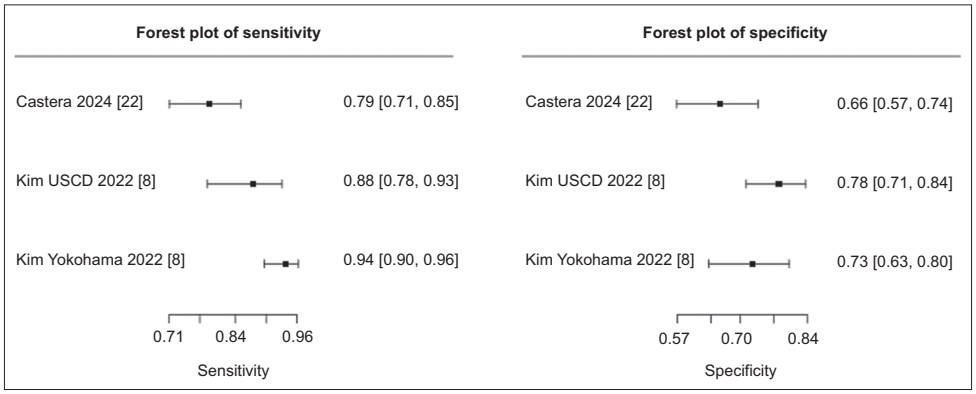
MEFIB (our meta-analysis)	
Sensitivity (rule out cut off)	0.77
Specificity (rule in cut off)	0.94
Grey zone magnitude	26.5%
FAST score	
Sensitivity (rule out cut off, <0.35)	0.89
Specificity (rule in cut off > 0.67)	0.89
Grey zone magnitude	33%
MAST score	
Sensitivity (rule out cut off, <0.165)	0.89
Specificity (rule in cut off, >0.242)	0.90
Grey zone magnitude	18.1%

Data for this table were obtained by Ravaioli et al (PMID: 36599683) and Noureddin et al (PMID: 34798176)

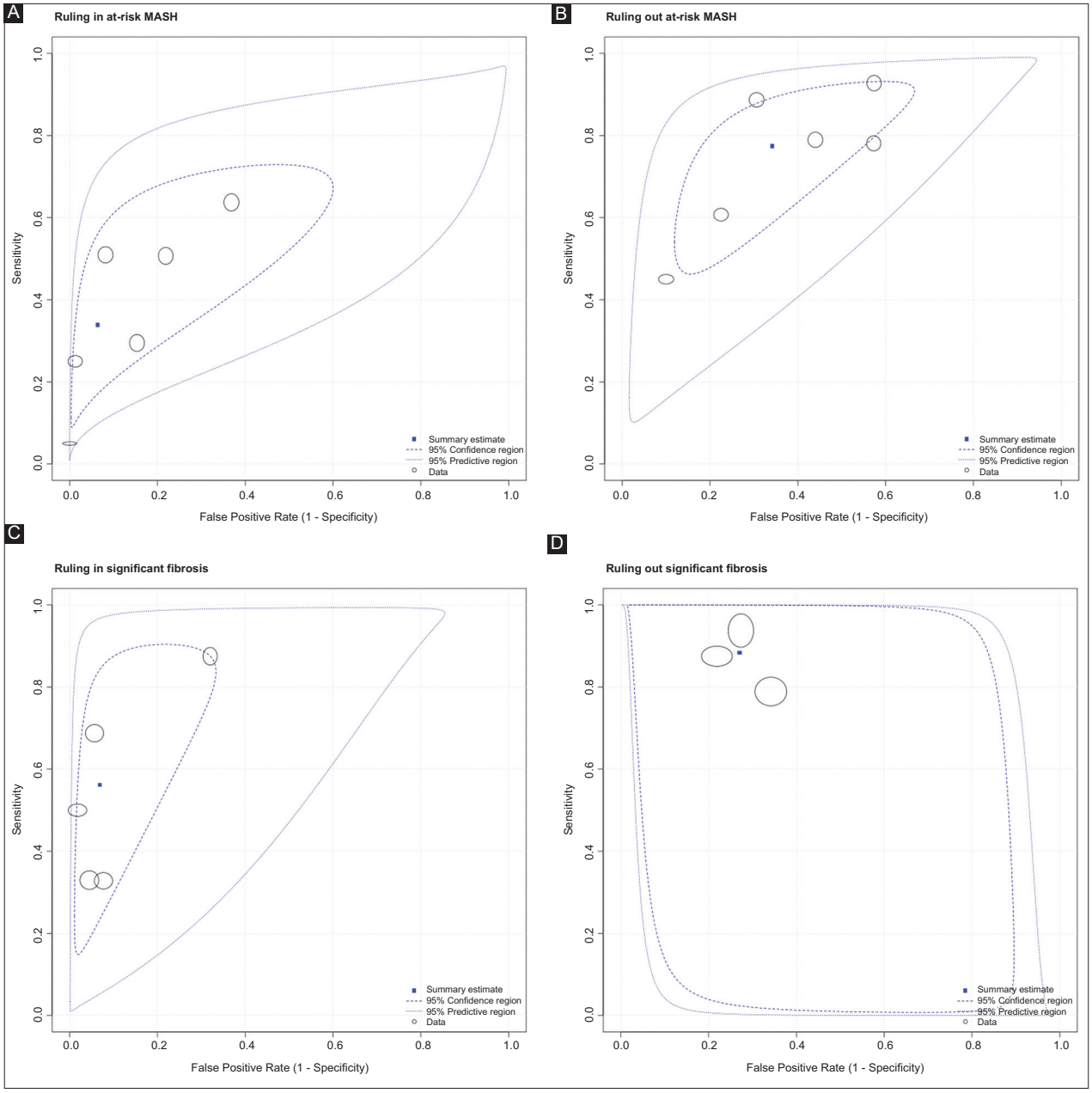
MEFIB, magnetic resonance elastography combined with the fibrosis-4 index; FAST, FibroScan-AST score; MAST, MRI-AST score; MASH, metabolic dysfunction-associated steatohepatitis



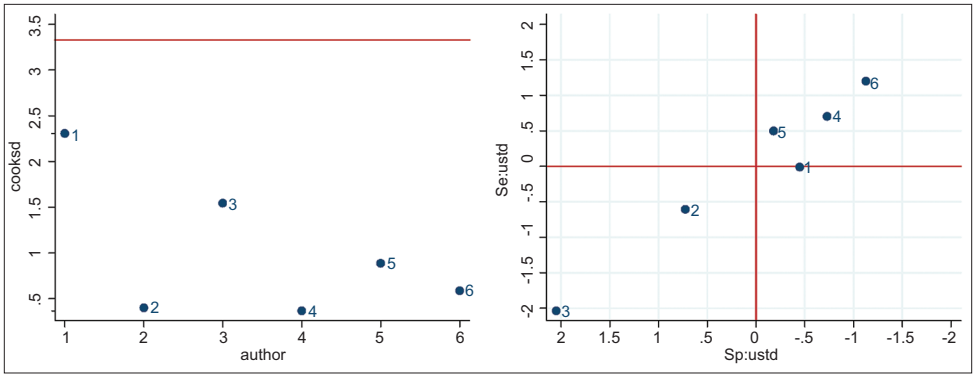
Supplementary Figure 1 Coupled forest plot of sensitivity and specificity of MEFIB index for ruling out at-risk MASH
MEFIB, magnetic resonance elastography combined with the fibrosis-4 index; MASH, metabolic dysfunction-associated steatohepatitis; CI, confidence interval



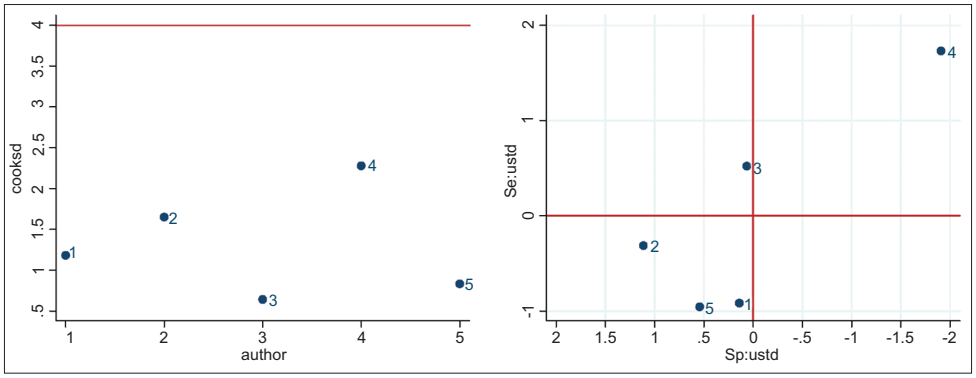
Supplementary Figure 2 Coupled forest plot of sensitivity and specificity of MEFIB index for ruling out significant fibrosis
MEFIB, magnetic resonance elastography combined with the fibrosis-4 index



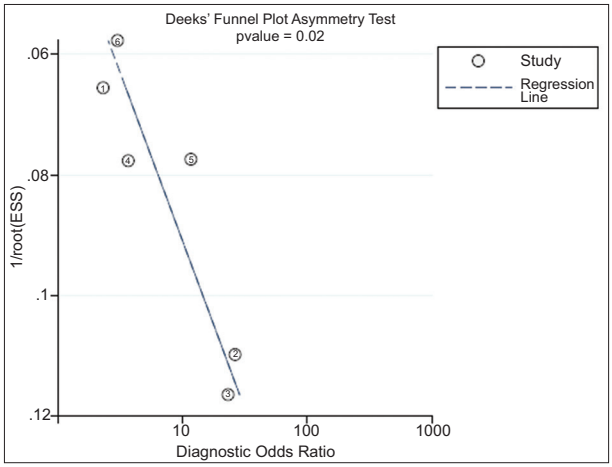
Supplementary Figure 3 (A-D) Pooled and individual study estimates of MEFIB index in the receiver-operating characteristic (ROC) space
MEFIB index, magnetic resonance elastography combined with the fibrosis-4 index; MASH, metabolic dysfunction-associated steatohepatitis



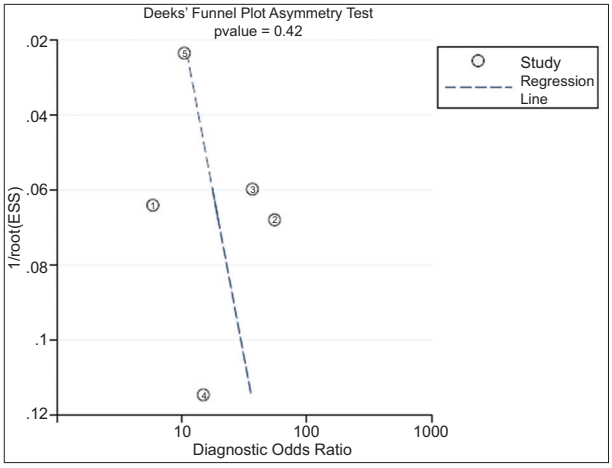
Supplementary Figure 4 Influence analysis for ruling in at-risk MASH
Left panel: Cook's distance. Cutoff for declaring Cook's distance to be large = 3.3 (red line). Right panel: standardized residuals (standardized predicted random effects). ustd, standardized residuals. 1=Castera 2024, 2= Qi 2024, 3= Noureddin 2023, 4= Imajo 2023, 5= Kim USCD Cohort 2022, 6= Kim Yokohama Cohort 2022
MASH, metabolic dysfunction-associated steatohepatitis



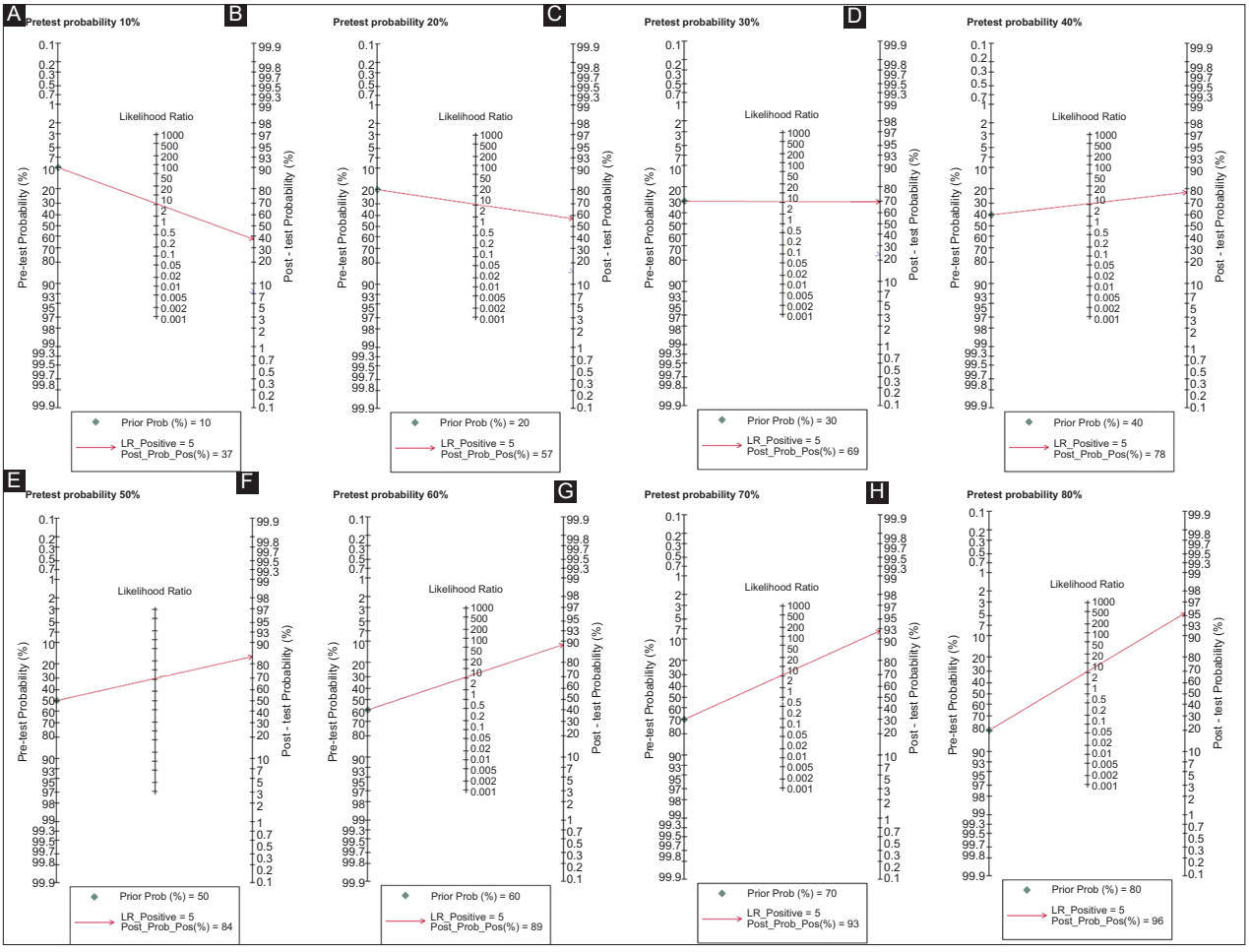
Supplementary Figure 5 Influence analysis for ruling in significant fibrosis
Left panel: Cook's distance. Cutoff for declaring Cook's distance to be large = 4 (red line). Right panel: standardized residuals (standardized predicted random effects). ustd, standardized residuals. 1=Castera 2024, 2= Kim USCD Cohort 2022, 3= Kim Yokohama Cohort 2022, 4= Inada 2022, 5= Loomba 2022



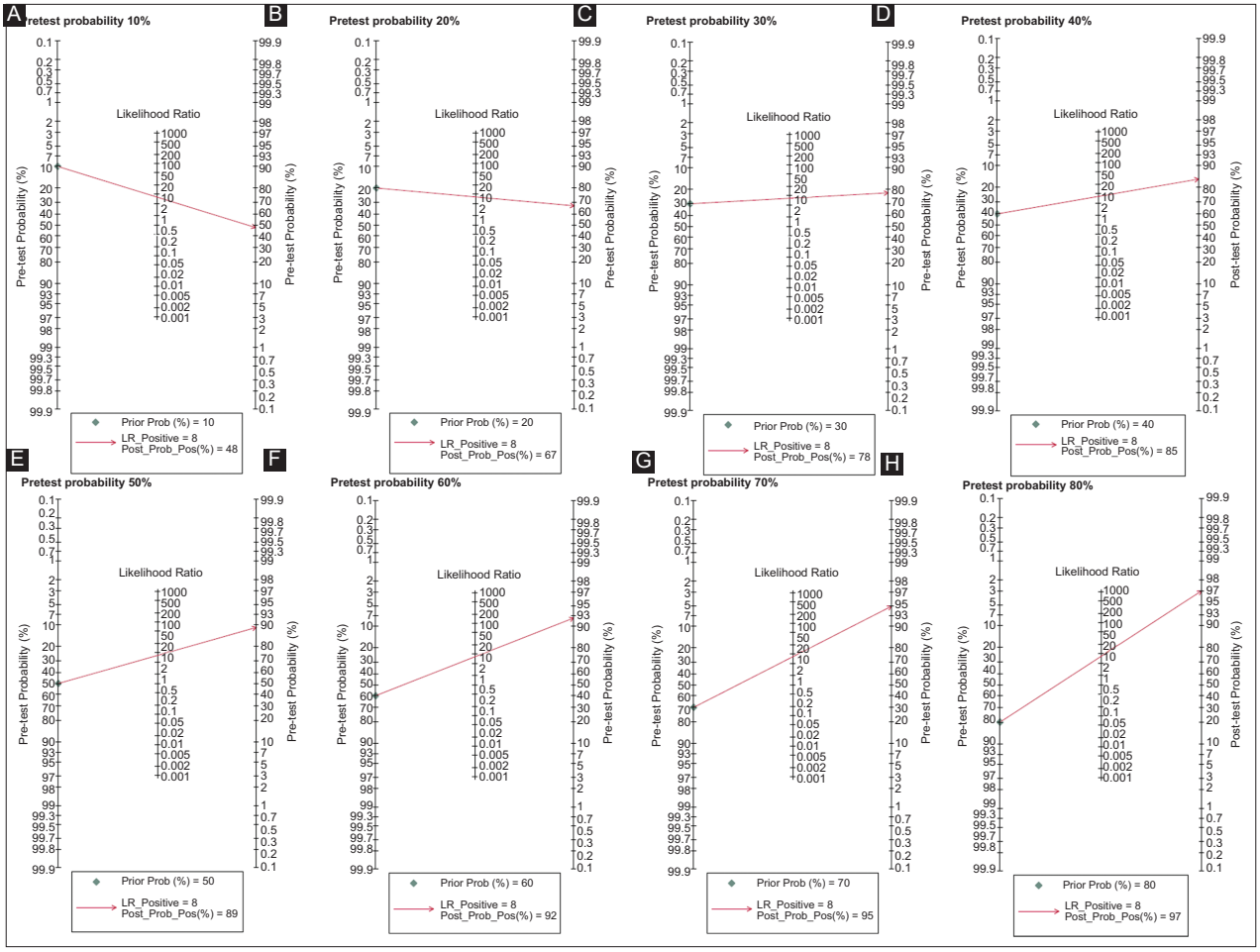
Supplementary Figure 6 Deeks' funnel plot for ruling in at risk MASH
MASH, metabolic dysfunction-associated steatohepatitis



Supplementary Figure 7 Deeks' funnel plot for ruling in significant fibrosis
MASH, metabolic dysfunction-associated steatohepatitis



Supplementary Figure 8 (A-H) Fagan nomograms for ruling in at risk MASH
MASH, metabolic dysfunction-associated steatohepatitis



Supplementary Figure 9 (A-H) Fagan nomograms for ruling in significant fibrosis