

Diagnostic Accuracy of Non-Invasive Diagnostic Tests for Nonalcoholic Fatty Liver Disease: A Systematic Review and Network Meta-Analysis

Yuxin Sun^{1,*}, Die Hu^{1,*}, Mingkun Yu², Shi-Bing Liang^{3–5}, Youyou Zheng⁶, Xin Wang⁷, Guangdong Tong^{1,8}

¹Shenzhen Traditional Chinese Medicine Hospital Affiliated to Nanjing University of Chinese Medicine, Shenzhen, People's Republic of China; ²Department of Oncology, Binzhou Hospital of Traditional Chinese Medicine, Binzhou, People's Republic of China; ³Clinical Study Center, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan, People's Republic of China; ⁴Centre for Evidence-Based Chinese Medicine, Shandong University of Traditional Chinese Medicine, Jinan, People's Republic of China; ⁵Postdoctoral Research Station, Shandong University of Traditional Chinese Medicine, Jinan, People's Republic of China; ⁶Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing, People's Republic of China; ⁷Department of Traditional Chinese Medicine, Sanbo Brian Hospital of Capital Medical University, Beijing, People's Republic of China; ⁸Department of Liver Disease, Shenzhen Traditional Chinese Medicine Hospital, Shenzhen, People's Republic of China

*These authors contributed equally to this work

Correspondence: Guangdong Tong, Shenzhen Traditional Chinese Medicine Hospital Affiliated to Nanjing University of Chinese Medicine, Shenzhen, People's Republic of China, Email tg755@163.com

Purpose: In recent decades, numerous non-invasive tests (NITs) for diagnosing nonalcoholic fatty liver disease (NAFLD) have been developed, however, a comprehensive comparison of their relative diagnostic accuracies is lacking. We aimed to assess and compare the diagnostic accuracy of various NITs for NAFLD using network meta-analysis (NMA).

Materials and Methods: We conducted a systematic search in seven databases up to April 2024 to identify studies evaluating the diagnostic values of NITs, with liver biopsy as the gold standard. The participants included patients with suspected or confirmed NAFLD, irrespective of age, sex, ethnicity. Statistical analysis was conducted using R 4.0.3 for Bayesian NMA and STATA 17.0 for pairwise meta-analysis. Sensitivity, specificity, diagnostic odds ratio (DOR), area under the receiver operating characteristic curve (AUC), and superiority index were calculated. Bayesian calculations were performed using the Rstan package, specifying parameters like MCMC chain count, iteration count, and operational cycles. The methodological quality of included studies was assessed using the QUADAS-2 tool.

Results: Out of 15,877 studies, 180 were included in the quantitative synthesis, and 102 were used in head-to-head meta-analyses. For diagnosing steatosis stage 1, Hydrogen Magnetic Resonance Spectroscopy (H-MRS, DOR 15,745,657.6, 95% CI 17.2–1,014,063.59) proved to be the most accurate. For significant fibrosis, HRI leading (DOR 80.94, 95% CI 6.46–391.41), For advanced fibrosis, CK-18 showed the highest performance (DOR 102654.16, 95% CI 1.6–134,059.8). For high-risk NASH, Real-Time Elastography showing the highest performance (DOR 18.1, 95% CI 0.7–96.33). Meta-regression analyses suggested that variability in the diagnostic accuracy of NITs for NAFLD may result from differences in study design, thresholds, populations, and performance indicators.

Conclusion: We conducted a network meta-analysis to rank the accuracy of these tests. While some results are promising, not all NITs demonstrate substantial accuracy, highlighting the need for validation with larger datasets. Future research should concentrate on studying the thresholds of NITs and enhancing the clarity of methodological reporting.

Keywords: NAFLD, NIT, obesity, NMA, diagnostic

Introduction

Due to the implementation of effective vaccines and potent antiviral therapies, the causes and patterns of the prevalence and incidence of end-stage liver diseases (ESLDs, eg, cirrhosis and HCC) have changed. Nonalcoholic fatty liver disease (NAFLD) significantly contributes to the prevalence of ESIDs.^{1,2} A recent study in China involving 5,757,335 check-up

participants estimated that the prevalence of liver steatosis was 44.39%, while advanced fibrosis and cirrhosis were 2.85% and 0.87%, respectively.³ NAFLD poses a substantial threat to patient health and imposes a significant burden on healthcare systems and socio-economic structures. Early diagnosis and effective management of NAFLD are essential. The diagnosis of NAFLD has relied on liver tissue biopsy as the gold standard, which is invasive, carries risks, and is limited by sampling bias and subjective interpretation issues. With advancements in medical technology and growing patient comfort demands, non-invasive tests (NITs) have garnered attention and widespread application. Particularly in early-stage disease or asymptomatic phases, these NITs enhance diagnostic accuracy and reduce the need for traditional liver tissue biopsy, thereby minimizing patient discomfort and healthcare costs. Numerous NITs have been developed to diagnose NAFLD, but their clinical application has yet to be widely adopted due to several challenges. First, more long-term validation, large-sample studies, and longitudinal research are needed to confirm their clinical efficacy.⁴ Second, there is a lack of standardization and differences in threshold settings for NITs, with varying recommendations for the same indicators, such as FIB-4 scores,⁵ Guidelines recommended the cutoffs proposed of <1.3 as the lower cutoff to rule out fibrosis and >2.67 as the upper cutoff to rule in fibrosis.^{6–10} The AASLD recommends <1.45 as the lower cutoff and >3.25 as the upper cutoff.¹¹ Third, many NITs do not meet the generally accepted cost-effectiveness thresholds, limiting their use,¹² and fourth, because the pathophysiology of NAFLD is complex, involving lipid accumulation, inflammation, oxidative stress, and fibrosis, most NITs focus on assessing fibrosis or fat content, which limits their ability to evaluate the disease comprehensively.^{13,14} In addition, the implementation and routine use of NITs in NAFLD faces multiple barriers, including a lack of societal awareness about NAFLD, skill gaps in interpreting test results, insufficient skills in using NITs to screen patients for timely referral, a misperception that NITs are not helpful in the absence of medical treatments, and limited predictive value for certain NITs.¹⁵ The diagnostic accuracy and target populations of different NITs still need to be clarified, and there is a lack of consensus in guidelines and expert opinions, which further complicates clinicians' decision-making when selecting appropriate diagnostic tools.¹⁶ Therefore, it is essential to clarify the diagnostic performance of NITs, optimize their ranking, and explore their potential value in various clinical contexts. This study assesses the sensitivity, specificity, and overall diagnostic value of existing NITs for NAFLD diagnosis through a systematic review and network meta-analysis, ranking their performance. The goal is to provide evidence-based support for clinical practice while identifying potential reasons for differences in diagnostic performance among NITs, offering insights for future improvements and optimization.

Materials and Methods

Protocol and Guidance

This study followed the Preferred Reporting Items for a Systematic Reviews and Meta-analysis of Diagnostic Test Accuracy (PRISMA-DTA) and Network Meta-Analyses (PRISMA-NMA),^{17,18} registered under PROSPERO (CRD42024582871).

Inclusion Criteria

We included diagnostic studies evaluating the predictive value of NITs in diagnosing NAFLD, with or without liver fibrosis or cirrhosis. The publication language was restricted to Chinese or English. Studies had to involve liver histopathology as the reference standard and adhere to established diagnostic criteria. The participants included patients with suspected or confirmed NAFLD, irrespective of age, sex, ethnicity. Combination with metabolic syndrome was allowed. NITs included both serum-based and imaging-based biomarkers. Studies were included if they reported data on true positive (TP), true negative (TN), false positive (FP), and false negative (FN), or metrics such as sensitivity (Se), specificity (Sp), accuracy, positive predictive value (PPV), negative predictive value (NPV), diagnostic odds ratio (DOR), and area under the receiver operating characteristic curve (AUC). While no restrictions were imposed on specific cut-off values for diagnostic measures such as Se, Sp, PPV, and NPV, we included all reported results with cut-off values to reflect the real-world variation in diagnostic thresholds. This approach aimed to capture the breadth of diagnostic performance across diverse clinical contexts. Variations in cut-off values and their potential impact on diagnostic

performance were explored qualitatively in the discussion. This manuscript covers numerous non-invasive diagnostic tests, the abbreviations for NITs can be found in [Supplemental eTable 1](#).

Exclusion Criteria

We excluded studies involving participants with fatty liver attributed to factors such as excessive alcohol consumption, genetic predispositions, or other secondary causes. Studies were excluded if they lacked full-text availability, were duplicate publications, or were conference abstracts without sufficient methodological details. To ensure the quality and reliability of included studies, we excluded those with unclear diagnostic criteria, sample sizes smaller than 30 participants, or insufficient data to calculate diagnostic accuracy measures. Only studies with observational designs (cross-sectional, case-control, or cohort) that specifically focused on the diagnostic performance of NITs were included.

Outcomes

The primary outcomes included absolute and relative Se and Sp, DOR, AUC and their respective 95% confidence intervals, which reflect the overall diagnostic accuracy of NITs in diagnosing NAFLD. Secondary outcomes included the superiority index (S index), results of subgroup analyses (eg, by disease severity or population characteristics). AUC values above 0.7 were considered indicative of acceptable diagnostic performance, while absolute Se and Sp values ≥ 0.8 were deemed clinically meaningful. To handle different reporting formats, we extracted or converted available data to calculate true positives, false positives, true negatives, and false negatives, ensuring consistent metric derivation across studies.

Search Strategy

We searched seven databases (PubMed, Embase, Cochrane Library, SinoMed, CNKI, VIP, and WANGFANG) for diagnostic studies up to April 28, 2024. The search strategy, including specific search terms, Medical Subject Headings (MeSH), and Boolean operators, was detailed in [Supplementary eTable 2](#). In addition, we performed manual searches of reference lists from relevant articles to identify additional studies. Efforts were also made to include grey literature by reviewing conference proceedings, dissertations, and unpublished data when available, aiming to minimize publication bias. Published journal articles and grey literature were both considered for inclusion in the analysis if they met the eligibility criteria.

Study Selection and Data Extraction

Two investigators independently screened the study titles and abstracts to identify studies meeting the criteria for full-text evaluation. Any disagreements at the title and abstract stage were discussed and resolved through consensus. Unresolved discrepancies were addressed by involving a senior third investigator for adjudication. All investigators underwent standardized training before the screening process, including calibration exercises, to ensure consistency in applying the inclusion and exclusion criteria. We used Endnote 21 software to record study selection decisions. Reasons for exclusion were systematically recorded at each stage (title/abstract and full-text screening) to ensure transparency and reproducibility. Diagnostic data (TP, FP, TN, FN), age, sex, disease condition, sample size, and research center information, were extracted. In cases where TP, TN, FP, or FN were not reported, they were calculated from known variables (Se, Sp, NPV, or PPV).

Quality Assessment and Publication Bias

The methodological quality of the included studies was assessed by two independent investigators using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) framework.¹⁹ The QUADAS-2 tool evaluates four key domains: patient selection, index test, reference standard, and flow and timing. Each domain was assessed for risk of bias, and the first three domains were also evaluated for concerns regarding applicability. Predefined signaling questions were used to guide judgments for each domain, with disagreements resolved through discussion or consultation with a third investigator. Quality scoring criteria were determined by categorizing studies as “low”, “moderate”, or “high” quality based on the level of risk of bias and applicability concerns identified in the QUADAS-2 assessment. Studies rated as having a high risk of bias in multiple domains were classified as low-quality. Low-quality studies were included in the primary analysis to ensure a comprehensive synthesis of the available evidence. However, their potential impact on

the robustness of the results was systematically addressed. Publication bias was assessed using funnel plots in STATA 17.0.

Data Synthesis

In the pairwise meta-analysis, we assessed Se, Sp, PPV, NPV, DOR, AUC and corresponding 95% confidence intervals (CIs). Head-to-head analyses were conducted only when more than ten studies were included. The Cochran's Q test and the inconsistency index (I^2) test were used to assess statistical heterogeneity. When little to no heterogeneity was present between studies ($I^2 < 25\%$), a fixed-effect model was used for data pooling. If substantial heterogeneity was present ($25\% < I^2 < 95\%$) and clinical heterogeneity was deemed acceptable, we applied a random-effects model. $I^2 > 50\%$, the meta-regression or sensitivity analysis were conducted to explore the sources of heterogeneity. If $I^2 > 95\%$ or if significant clinical heterogeneity was present, we refrained from pooling quantitative data. The funnel plot method was used to determine publication bias. The pairwise meta-analysis and meta-regression were performed using STATA software, version 17.0. Only studies with complete data or data that could be manually calculated were included, no missing data were present in this work.

The network meta-analysis was conducted with Bayesian framework using R 4.0.3. Data preprocessing included calculating true positives, true negatives, patients correctly diagnosed by the gold standard, and those excluded. Diagnostic test names were replaced with numeric codes. Preprocessed data were analyzed using an ANOVA random-effects model.^{20,21} The model assumed a common-effect parameter across comparisons for diagnostic odds ratios (DOR). Rankings based on DOR values indicated the diagnostic test's discrimination capability, with higher values indicating better discriminatory test performance.²² The primary outcomes included DOR, sensitivity, specificity, AUC, and 95% CIs. Bayesian calculations were performed using the Rstan package, specifying parameters like MCMC chain count, iteration count, and operational cycles (initially set at four chains, 50,000 iterations, with a 5-interval random-effects model to mitigate initial value impact). Default non-informative priors provided by the Rstan package were used for the Bayesian analysis, ensuring minimal prior influence on the parameter estimates. Model convergence was assessed using the potential scale reduction factor (R-hat). R-hat values below 1.1 were considered indicative of good convergence. Any convergence issues identified were resolved by increasing the number of iterations or adjusting model priors to improve stability. Network plots and inconsistency evaluation for NMA were not conducted, because all NITs were compared to liver biopsies, Closed-loop structures were failed to establish, which are essential for generating network plots and evaluating the consistency between direct and indirect evidence, were therefore absent.

Subgroup Analyses

For the different stages of NAFLD, we performed several subgroup analyses based on the steatosis stage, fibrosis stage, NAFLD with fibrosis, NASH, and high-risk NASH. We defined "significant fibrosis" as $\geq S2$ on the Scheuer and METAVIR scales or $\geq S3$ on the Ishak scale, and "advanced fibrosis" as $\geq S3$ on the Scheuer and METAVIR scales or $S4$ on the Ishak scale.^{23–26} Meta-regression was conducted to explore potential sources of heterogeneity, considering factors such as sample size, publication year, number of centers, funding, and the four QUADAS-2 domains: patient selection, index test, reference standard, and flow and timing.

Results

Eligible Studies and Study Characteristics

We identified 15,877 studies, and after removing duplicates, 4,213 studies underwent screening (Figure 1). A full-text review of 738 studies resulted in the inclusion of 180 studies evaluating 153 NITs. All studies utilized liver biopsy as the gold standard comparator. Geographically, most studies were conducted in China (62 studies) and the USA (33 studies). The characteristics and reference citations of the included studies are detailed in Supplemental eTable 3. To enhance clarity, NITs were categorized into serological biomarker models and imaging-based tests. Among the serological models, the FIB4 index was the most studied (52 studies, 64 arms), followed by the NAFLD Fibrosis Score (NFS, 39 studies, 47 arms), the APRI (34 studies, 45 arms). For imaging-based tests, transient elastography (TE) had the most

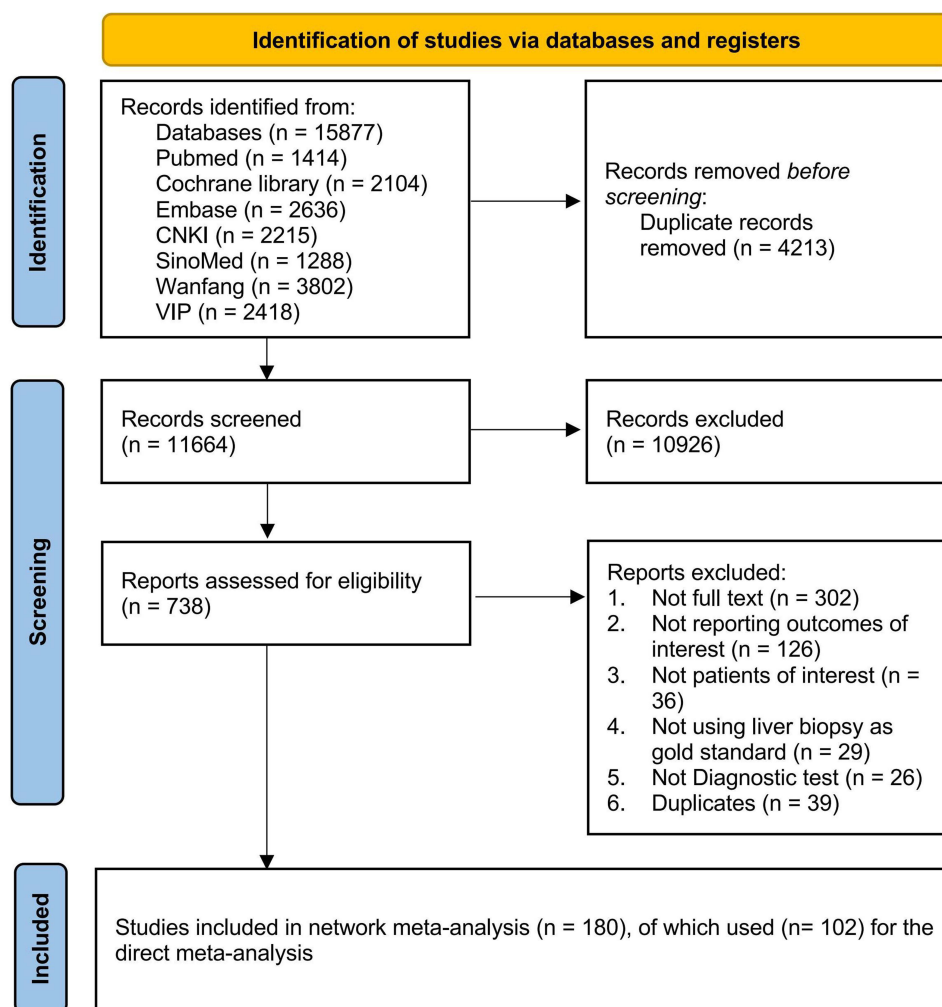


Figure 1 Flow diagram showing the inclusion of studies from the literature review.

studies (43 studies, 100 arms), followed by PDFF (12 studies, 18 arms), MRE (11 studies, 19 arms). Details shown in [Figure 2](#). Additionally, 94 NITs were evaluated in only one study each.

Pairwise Meta-Analysis for DTA

A direct meta-analysis was conducted to evaluate the diagnostic accuracy of 11 NITs for NAFLD across various disease stages. The analysis included data from 101 studies and focused on detecting steatosis ([S1](#) and [S2](#)), significant fibrosis (\geq F2), advanced fibrosis (\geq F3), and high-risk NASH. Imaging-based markers, such as MRE, TE, ARFI, and PDFF, generally outperformed serum-based markers in diagnosing fibrosis stages, while simpler indices (APRI, AAR, and BARD) showed lower diagnostic accuracy. Details see [Table 1](#). No significant publication bias was identified ([Figure 3](#)). SROC curve see [Figure 4](#), forest plot see [Figure 5](#).

For advanced fibrosis (\geq F3), MRE achieved the highest accuracy (AUC 0.93, 95% CI 0.90–0.95), with Se 0.86 (95% CI 0.80–0.90), Sp 0.91 (95% CI 0.86–0.94), and DOR 60 (95% CI 31–119). Among serum markers, FIB-4 (AUC 0.84, 95% CI 0.81–0.87) and NFS (AUC 0.82, 95% CI 0.78–0.85) performed better than APRI (AUC 0.79, 95% CI 0.75–0.82). For significant fibrosis (\geq F2), imaging tests, MRE showed strong diagnostic utility with DOR values above 31 and AUC 0.92. Serum markers, such as FIB-4 (AUC 0.80, 95% CI 0.76–0.83) and PRO-C3 (AUC 0.83, 95% CI 0.80–0.86), exhibited moderate performance. For steatosis (S1 and S2), PDFF achieved excellent diagnostic accuracy with an AUC of 0.95 (95% CI 0.93–0.96) for S1 and 0.89 (95% CI 0.86–0.92) for S2. TE also demonstrated high

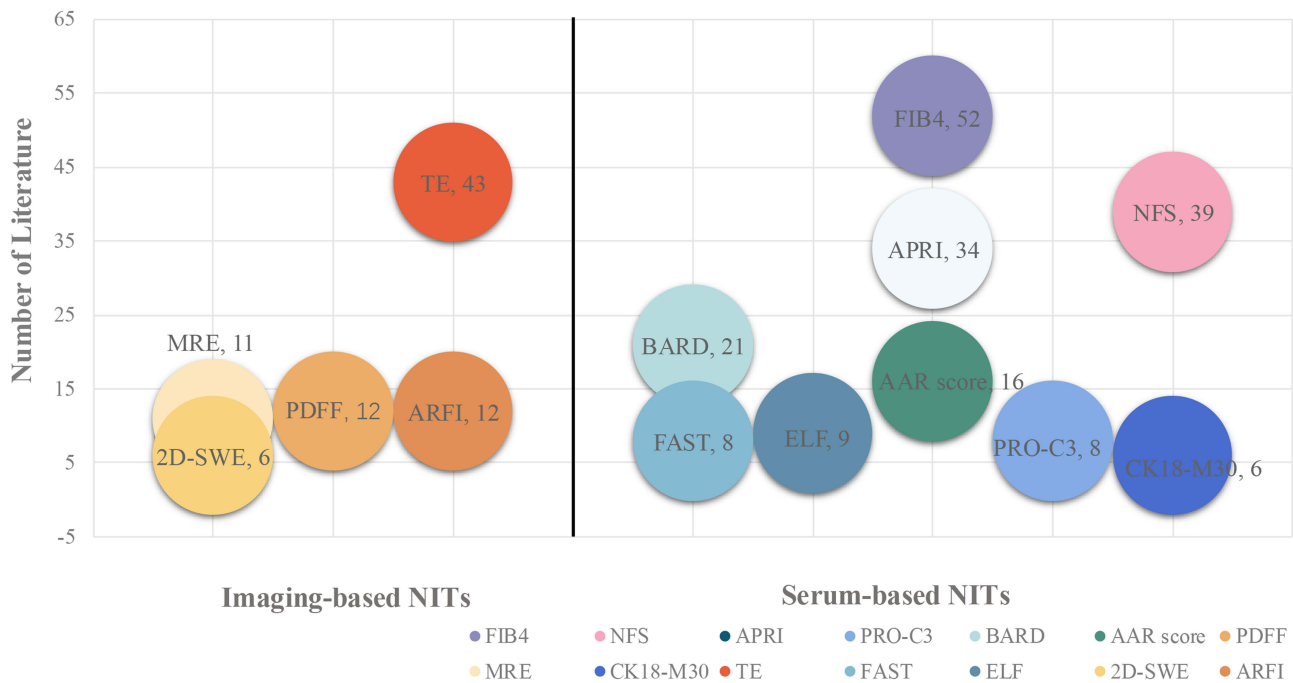


Figure 2 Evidence Map of NITs (studies or multi-arm > 10) in the prognostic of NAFLD. Evidence map displaying the distribution of NITs based on their modality and literature size. The X-axis distinguishes between imaging-based NITs (left) and serum-based NITs (right). The Y-axis represents the number of the literature.

accuracy for S1 with an AUC of 0.92 (95% CI 0.89–0.94). Meta-regression analyses suggested that the heterogeneity in the diagnostic accuracy of NITs for NAFLD may be due to differences in study design, threshold settings, study populations, and performance indicators across studies. Details in Figure 6.

Table 1 Pairwise Meta-Analysis of Diagnostic Accuracy for NAFLD

Subgroups	NITs	Sen (95% CI)	Spe (95% CI)	DOR (95% CI)	AUC (95% CI)
NAFLD ≥F3	AAR	0.60[0.50, 0.69]	0.70[0.61, 0.77]	3[2, 6]	0.70[0.65–0.73]
	APRI	0.59[0.47, 0.87]	0.82[0.74, 0.87]	6[4, 10]	0.79[0.75–0.82]
	BARD	0.69[0.60, 0.76]	0.67[0.60, 0.73]	4[3, 6]	0.73[0.69–0.77]
	FIB4	0.63[0.52, 0.74]	0.87[0.81, 0.92]	12[8, 17]	0.84[0.81–0.87]
	NFS	0.66[0.53, 0.78]	0.82[0.73, 0.88]	9[5, 14]	0.82[0.78–0.85]
	PRO-C3	0.71[0.61, 0.79]	0.74[0.66, 0.80]	7[5, 9]	0.79[0.75, 0.82]
	MRE	0.86[0.80, 0.90]	0.91[0.86, 0.94]	60[31, 119]	0.93[0.90, 0.95]
	TE	0.88[0.83, 0.91]	0.80[0.75, 0.84]	28[18, 45]	0.91[0.88, 0.93]
	ARFI	0.83[0.76, 0.88]	0.87[0.81, 0.91]	32[22, 47]	0.92[0.89, 0.94]
	NAFLD ≥F2	APRI	0.73[0.66, 0.79]	0.66[0.59, 0.73]	5[4, 7]
NFS		0.69[0.58, 0.78]	0.75[0.66, 0.83]	7[4, 12]	0.79[0.75–0.82]
FIB4		0.64[0.52, 0.75]	0.83[0.72, 0.91]	9[5, 15]	0.80[0.76, 0.83]
MRE		0.84[0.76, 0.90]	0.86[0.79, 0.91]	31[21, 46]	0.92[0.89, 0.94]
PRO-C3		0.59[0.48, 0.68]	0.85[0.80, 0.89]	8[6, 12]	0.83[0.80, 0.86]
TE		0.76[0.67, 0.83]	0.80[0.75, 0.84]	12[9, 17]	0.85[0.81, 0.88]
ARFI		0.73[0.61, 0.82]	0.82[0.69, 0.90]	12[7, 20]	0.83[0.80, 0.86]
NASH vs NAFL	CK18-M30	0.67[0.54, 0.78]	0.90[0.80, 0.96]	18[9, 38]	0.86[0.83–0.89]
Steatosis (≥S1)	TE	0.85[0.80, 0.89]	0.85[0.79, 0.89]	31[17, 56]	0.92[0.89, 0.94]
	PDFF	0.92[0.88, 0.94]	0.86[0.75, 0.92]	66[27, 157]	0.95[0.93, 0.96]
Steatosis (≥S2)	PDFF	0.88[0.82, 0.93]	0.88[0.82, 0.92]	55[27, 112]	0.89[0.86, 0.92]

Notes: Reference citations included in this table could be found in [supplementary table 4](#).

Abbreviation: List of abbreviations were provided in [Supplemental eTable 1](#).

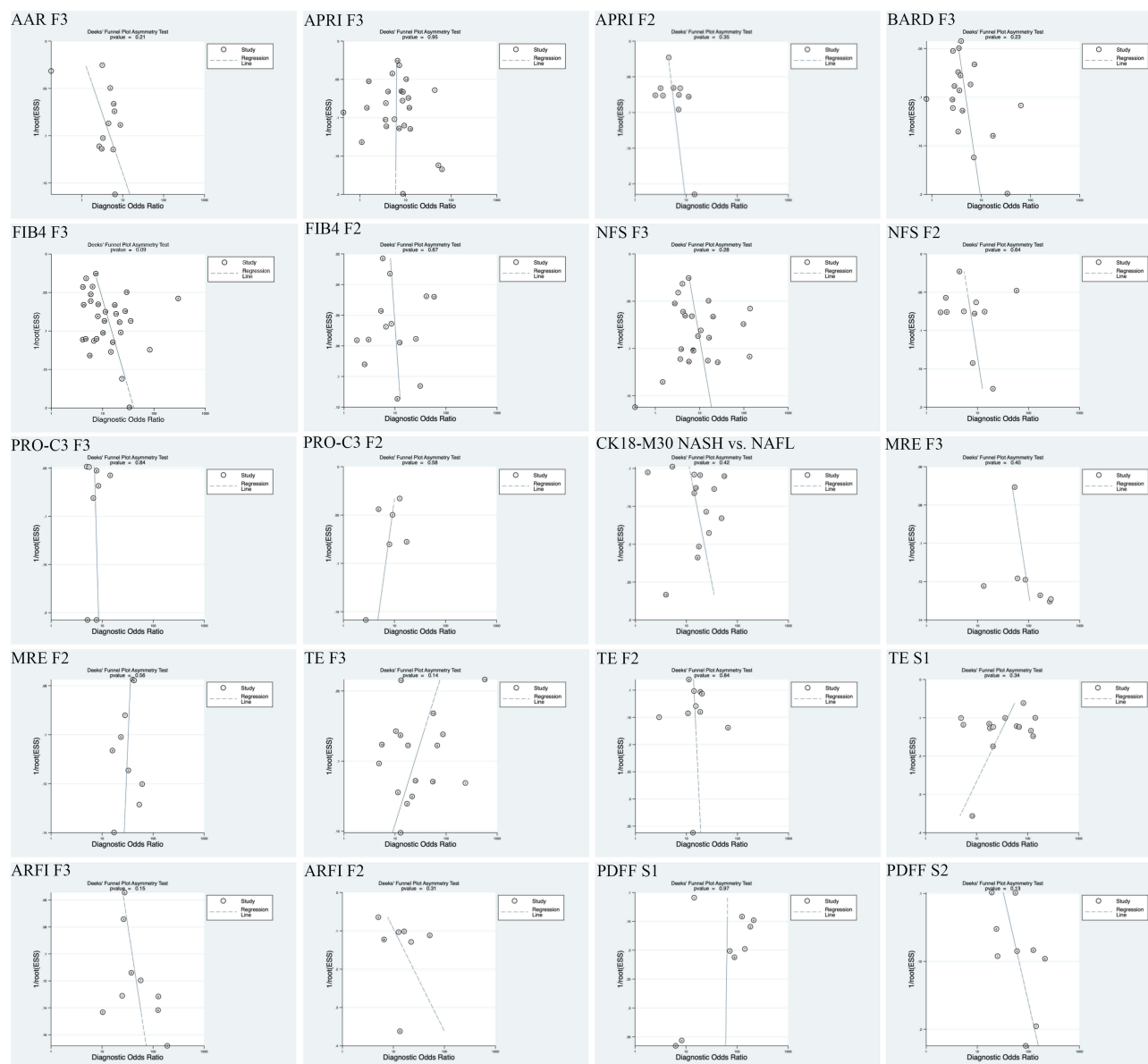


Figure 3 Deeks' Funnel plots for diagnostic comparisons.

Abbreviations: AAR, Aspartate Aminotransferase to Alanine Aminotransferase Ratio score; APRI, Aspartate Aminotransferase to Platelet Ratio Index; F3, Fibrosis Stage 3; F2, Fibrosis Stage 2; NFS, NAFLD Fibrosis Score; MRE, Magnetic Resonance Elastography; TE, Transient Elastography; ARFI, Acoustic Radiation Force Impulse Imaging; PDFF, Proton Density Fat Fraction.

Network Meta-Analysis for DTA

A total of 180 studies were available for NMA. Seven subgroup analyses were conducted using disease stage as the grouping criterion, including steatosis \geq S1, \geq S2, NAFL vs NASH, NAFLD with fibrosis, NAFLD \geq F3, NAFLD \geq F2, and high-risk NASH.

For steatosis (\geq S1), H-MRS achieved the highest diagnostic accuracy (DOR 15,745,657.6, 95% CI 17.2–1,014,063.59), followed by NLV (DOR 515,080.28, 95% CI 3.05–295,953.65) and among serum markers, ALT exhibited the best performance (DOR 722.13, 95% CI 7.48–3,133.23), followed by TyG (DOR 241.24, 95% CI 2.56–1,229.06). **For steatosis (\geq S2),** HRI (DOR 127.23, 95% CI 7.69–594.44) and MRS (DOR 116.52, 95% CI 5.21–595.46) outperformed PDFF (DOR 28.88, 95% CI 11.98–56.75) and TE (DOR 10.7, 95% CI 3.72–22.69). Details are provided in [Table 2](#) and [3](#).

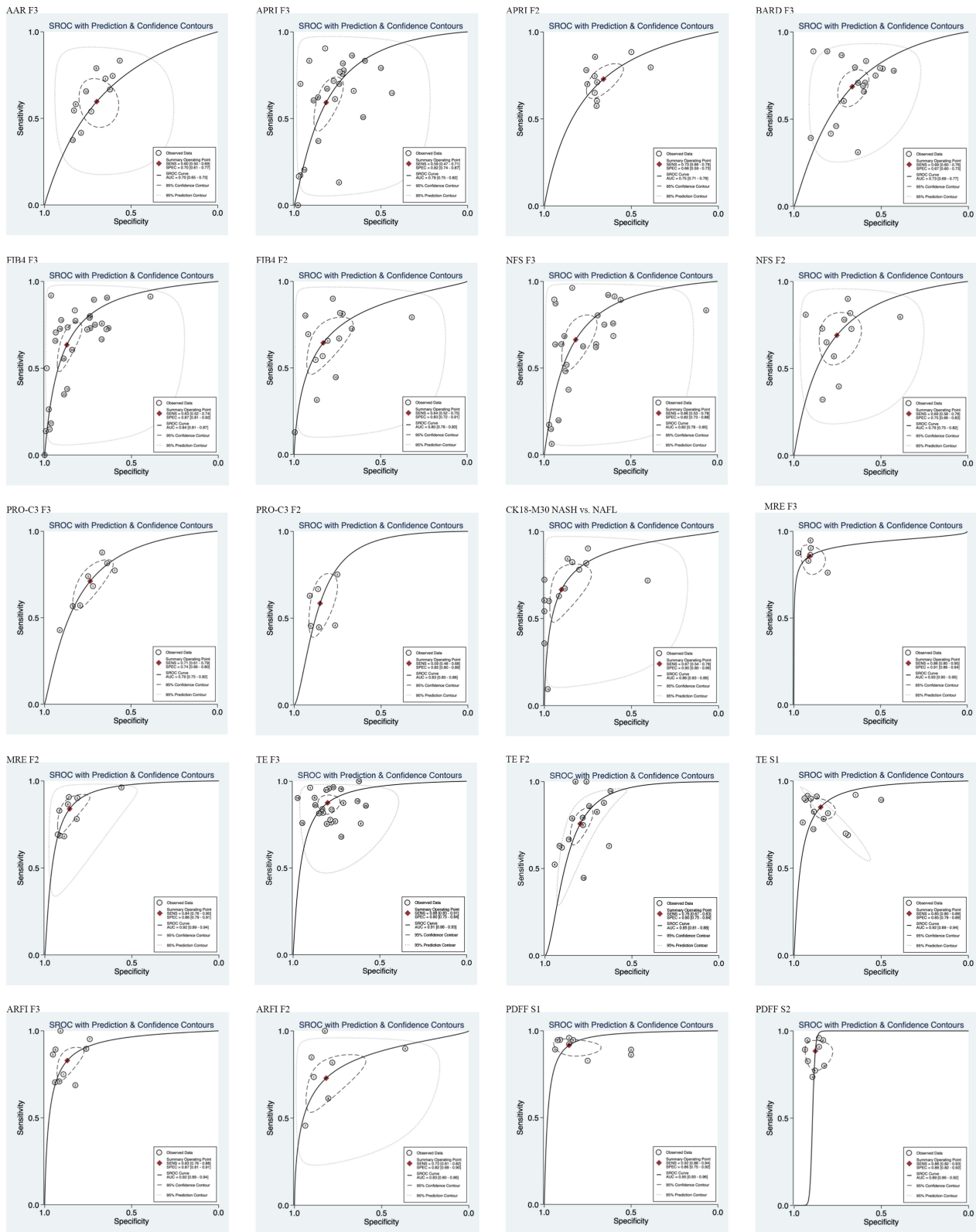


Figure 4 SROC Curve Comparing Various NITs for Diagnosing NAFLD in Direct Comparison.

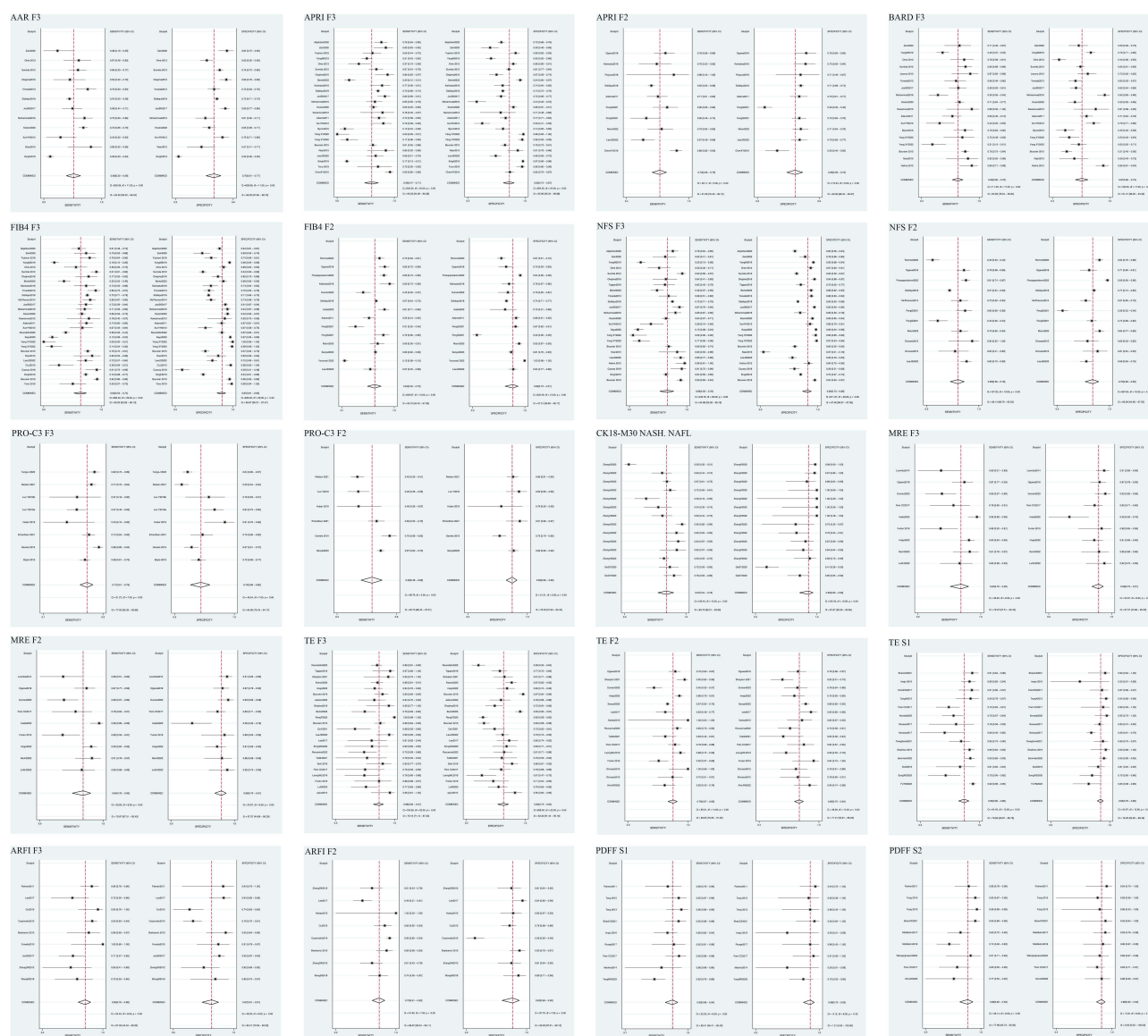


Figure 5 Forest Plot Comparing Various NITs for Diagnosing NAFLD in Direct Comparison.

For distinguishing NAFLD and NASH, MRE showed the highest performance (DOR 4360974.03, 95% CI 210.74–3,589,007.06), followed by HsCRP (DOR 942765.03, 95% CI 4.88–154,229.3), IL-1RA (DOR 1234.67, 95% CI 10.55–7459.82). Other NITs, such as miRNA-99a, A1AT, PC, FNI, SAN, NI-NASH-DS, demonstrated moderate accuracy (DORs ranging from 282.72 to 42.95). Details are provided in [Table 4](#).

For NAFLD with any stage of fibrosis, sH2a showed the best performance (DOR 40.64, 95% CI 0.64, 244.34), followed by IgA (DOR 27.38, 95% CI 0.78, 157.05), sICAM-1 (DOR 23.18, 95% CI 0.45, 152.53), AAR score (DOR 18.24, 95% CI 0.33–108.01), NIS (DOR 10.13, 95% CI 0.34–59.72), and TE (DOR 9.98, 95% CI 1.7–32.59). Details are shown in [Table 5](#).

For significant fibrosis (\geq F2), imaging-based NITs demonstrated the highest diagnostic accuracy, with HRI leading (DOR 80.94, 95% CI 6.46–391.41), followed by MRE (DOR 22.36, 95% CI 13.17–36.15) and 2D-SWE (DOR 21.72, 95% CI 5.59–61.32). Among serum-based NITs, HSI showed the best performance (DOR 63.75, 95% CI 13.13–196.85), followed by SHG B-index (DOR 40.57, 95% CI 4.84–160.84), FLI (DOR 30.75, 95% CI 6.37–93.26), Ali2021 (DOR 28.01, 95% CI 4.85–89.58), LRM (DOR 25.45, 95% CI 6.99–64.38). Details are shown in [Table 6](#).

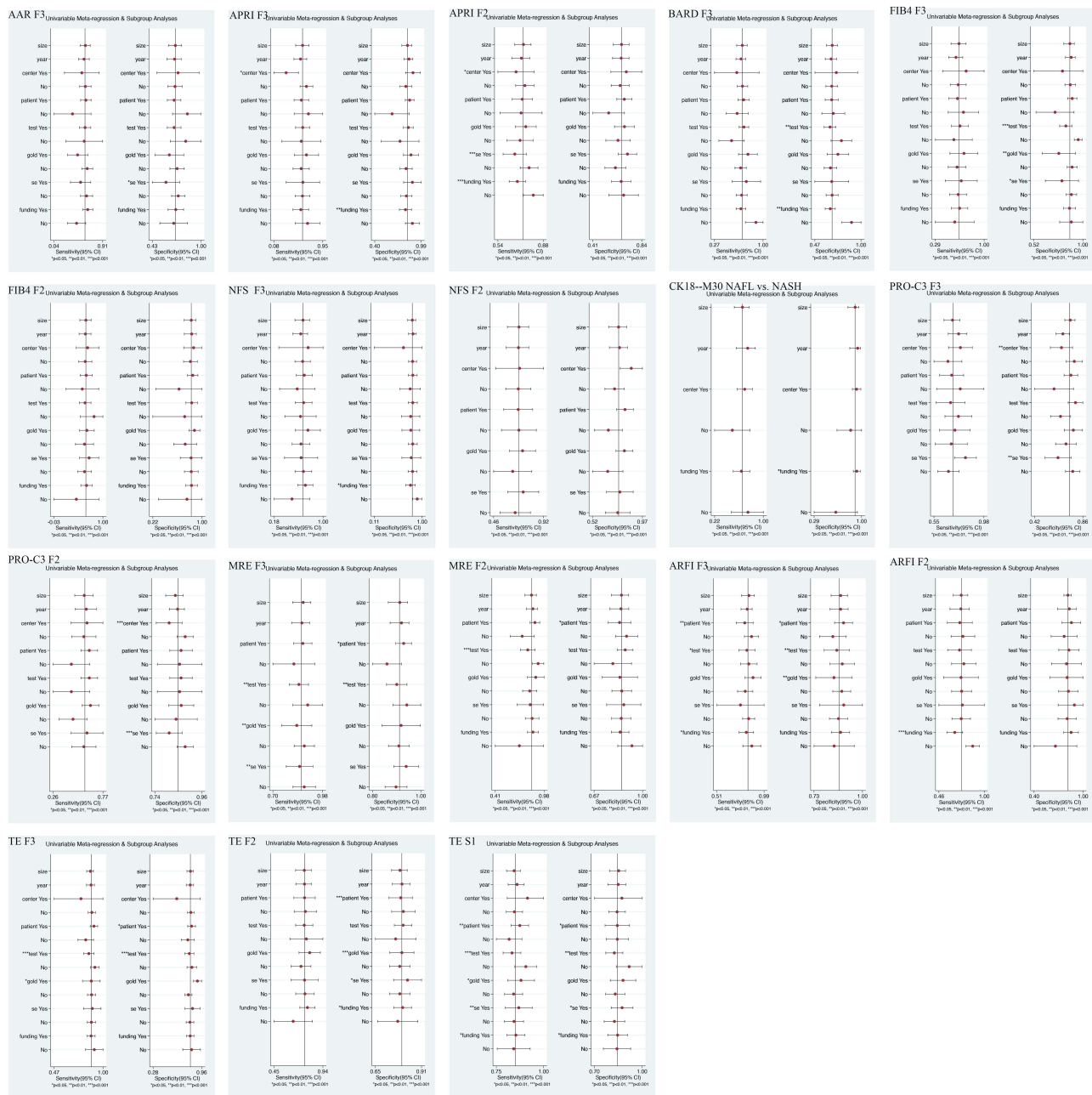


Figure 6 Meta-regression in Direct Comparison of NITs for Diagnosing NAFLD. The meta-regression analysis includes the following covariates: Size (sample size), Year (year of paper publication), Patient (selection bias in patient inclusion), Gold (implementation of the gold standard), Se (sensitivity), Center (number of research centers), Test (bias in test implementation or interpretation), and Funding (research funding sources).

Abbreviations: AAR, Aspartate Aminotransferase to Alanine Aminotransferase Ratio score; APRI, Aspartate Aminotransferase to Platelet Ratio Index; ARFI, Acoustic Radiation Force Impulse Imaging; F3, Fibrosis Stage 3; F2, Fibrosis Stage 2; MRE, Magnetic Resonance Elastography; NFS, NAFLD Fibrosis Score; S1, Steatosis Stage 1; TE, Transient Elastography; 95% CI, 95% Confidence Interval.

For advanced fibrosis (\geq F3), serum-based NITs showed the highest diagnostic accuracy, with CK-18 leading (DOR 102654.16, 95% CI 1.6–134,059.8), followed by SHG B-index (DOR 1149.9, 95% CI 11.35–5983.64), LFS (DOR 793.48, 95% CI 4.17–5316.76), LPR (DOR 239.87, 95% CI 3.96–1371.41), FM (DOR 78.67, 95% CI 4.6–382.09), HSI (DOR 61.12, 95% CI 3.53–306.22), FLI (DOR 52.18, 95% CI 3.43–247.93), 10 metabolites (DOR 50.89, 95% CI 1.88–271.77), and Hepascore (DOR 47.17, 95% CI 10.81–141.89). Among imaging-based NITs, MRE had a DOR of 51.11 (95% CI 18.71–115.19). Details are shown in [Table 7](#).

Table 2 27 NITs Diagnostic for Steatosis \geq S1 ANOVA Model Analysis Results

No.	NITs	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	S index (95% CI)	Relative Sensitivity (95% CI)	Relative Specificity (95% CI)
1	H-MRS	0.81(0.59, 0.95)	0.98(0.83, 1)	15,745,657.6(17.2, 1,014,063.59)	21.36(0.52, 49)	1	1
2	NLV	0.74(0.4, 0.95)	0.93(0.57, 1)	515,080.28(3.05, 295,953.65)	12.36(0.04, 47)	0.93(0.48, 1.37)	0.96(0.58, 1.15)
3	NLV-SD	0.72(0.38, 0.93)	0.93(0.57, 1)	325,159.91(2.71, 395,626.93)	10.23(0.03, 45)	0.9(0.45, 1.36)	0.96(0.59, 1.15)
4	ALT	0.84(0.59, 0.97)	0.91(0.65, 1)	722.13(7.48, 3133.23)	16.59(0.16, 47)	1.06(0.7, 1.5)	0.94(0.66, 1.13)
5	TyG	0.74(0.42, 0.93)	0.88(0.53, 1)	241.24(2.56, 1229.06)	5.84(0.03, 39)	0.93(0.52, 1.36)	0.9(0.55, 1.1)
6	FAI	0.86(0.58, 0.97)	0.81(0.4, 0.99)	234.73(2.89, 1395.87)	11.39(0.04, 45)	1.07(0.71, 1.48)	0.83(0.41, 1.06)
7	VAI	0.74(0.45, 0.93)	0.88(0.54, 1)	190.06(2.48, 996.4)	5.33(0.03, 37.05)	0.92(0.52, 1.36)	0.9(0.56, 1.09)
8	CK18-M30	0.65(0.4, 0.84)	0.93(0.75, 1)	175.38(4.36, 1065.58)	2.57(0.04, 19)	0.81(0.48, 1.2)	0.96(0.75, 1.15)
9	FGF21	0.76(0.43, 0.95)	0.81(0.42, 0.99)	158.92(1.56, 604.62)	5.13(0.03, 39)	0.95(0.52, 1.38)	0.84(0.42, 1.07)
10	PDFF	0.91(0.83, 0.96)	0.86(0.69, 0.96)	97.36(18.12, 330.99)	16.07(1.15, 41.05)	1.14(0.95, 1.55)	0.88(0.7, 1.05)
11	HSI	0.59(0.26, 0.86)	0.88(0.56, 1)	85.1(1.22, 508.25)	1.46(0.03, 11)	0.73(0.32, 1.18)	0.91(0.58, 1.11)
12	Liver Young's modulus	0.82(0.55, 0.96)	0.8(0.48, 0.97)	56.09(2.69, 282.84)	6.23(0.03, 37)	1.03(0.67, 1.42)	0.82(0.5, 1.05)
13	CK18-M65	0.9(0.7, 0.99)	0.57(0.15, 0.94)	53.03(1.15, 289.57)	6.12(0.03, 37)	1.12(0.82, 1.54)	0.59(0.15, 0.97)
14	VTQ	0.9(0.67, 0.98)	0.6(0.21, 0.92)	43.67(1.44, 233.66)	6.19(0.03, 33)	1.12(0.82, 1.55)	0.62(0.21, 0.96)
15	MRI	0.83(0.67, 0.92)	0.85(0.69, 0.94)	40.39(7.92, 123.22)	4.63(0.14, 31)	1.03(0.79, 1.43)	0.88(0.71, 1.04)
16	FLI	0.71(0.41, 0.92)	0.8(0.41, 0.98)	32.67(1.22, 177.32)	1.96(0.03, 19)	0.89(0.5, 1.31)	0.82(0.42, 1.04)
17	UGAP	0.81(0.54, 0.96)	0.7(0.33, 0.94)	24.84(1.56, 121.26)	2.82(0.03, 25)	1.01(0.64, 1.43)	0.72(0.34, 0.99)
18	UAP	0.78(0.56, 0.91)	0.81(0.58, 0.94)	24.33(3.46, 83.86)	2.15(0.03, 17)	0.97(0.66, 1.36)	0.83(0.59, 1.02)
19	NFS	0.62(0.3, 0.87)	0.8(0.4, 0.98)	21.37(0.79, 105.36)	0.83(0.02, 5.4)	0.77(0.35, 1.24)	0.82(0.41, 1.05)
20	GGT	0.72(0.4, 0.92)	0.75(0.39, 0.95)	20.45(0.93, 102.97)	1.57(0.02, 11.67)	0.9(0.47, 1.36)	0.77(0.4, 1.01)
21	AST	0.87(0.63, 0.98)	0.54(0.2, 0.85)	19.98(1.02, 100.73)	2.76(0.03, 21)	1.09(0.75, 1.52)	0.55(0.2, 0.88)
22	TE	0.79(0.73, 0.84)	0.8(0.73, 0.86)	16.7(8.99, 28.13)	0.93(0.12, 3)	0.99(0.81, 1.35)	0.82(0.74, 0.98)
23	ATI	0.78(0.48, 0.94)	0.64(0.28, 0.92)	14.77(0.93, 67.32)	1.47(0.02, 13)	0.98(0.58, 1.4)	0.66(0.29, 0.96)
24	CK-18	0.82(0.62, 0.94)	0.62(0.33, 0.85)	12.4(1.57, 49.3)	1.19(0.03, 8.33)	1.03(0.73, 1.45)	0.63(0.33, 0.89)
25	CTL-S	0.85(0.59, 0.97)	0.49(0.19, 0.8)	10.85(0.76, 49.28)	1.39(0.02, 13)	1.06(0.71, 1.48)	0.5(0.2, 0.84)
26	Clinical-CT index	0.85(0.6, 0.97)	0.45(0.15, 0.77)	9.31(0.63, 42.63)	1.1(0.02, 11)	1.06(0.7, 1.46)	0.46(0.16, 0.81)
27	SteatoTest	0.79(0.5, 0.95)	0.51(0.18, 0.83)	8.45(0.54, 40.74)	0.88(0.02, 7)	0.99(0.6, 1.42)	0.52(0.18, 0.87)

Notes: Reference citations included in this table could be found in the [supplementary table 5](#).

Abbreviation: List of abbreviations were provided in [Supplemental eTable 1](#).

For high-risk NASH (NAS \geq 4, \geq F2), imaging-based NITs demonstrated varying levels of diagnostic accuracy, with Real-Time Elastography showing the highest performance (DOR 18.1, 95% CI 0.7–96.33), followed by MRE (DOR 11.27, 95% CI 0.64–53.29) and TE (DOR 6.36, 95% CI 1.77–16.52). Among serum-based NITs, CK-18 achieved the most remarkable performance (DOR 39.46, 95% CI 0.48–267.09), followed by ELF (DOR 34.64, 95% CI 0.83–181.08), C-DAG (DOR 29.34, 95% CI 1.59–144.73), FNI (DOR 17.36, 95% CI 2.48–59.67), MACK-3 (DOR 10.63, 95% CI 1.82–38.35), MEFIB (DOR 8.34, 95% CI 1.47–27.15), NIS (DOR 7.9, 95% CI 1.62–23.06), FAST (DOR 4.67, 95% CI 2.27–8.35), APRI (DOR 3.86, 95% CI 1.27–9), FIB4 (DOR 3.78, 95% CI 1.57–7.51), and NFS (DOR 2.93, 95% CI 0.97–6.55). Details are shown in [Table 8](#).

Table 3 4 NITs Diagnostic for Steatosis \geq S2 ANOVA Model Analysis Results

No.	NITs	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	S index (95% CI)	Relative Sensitivity (95% CI)	Relative Specificity (95% CI)
1	HRI	0.86(0.63, 0.97)	0.87(0.65, 0.98)	127.23(7.69, 594.44)	3.56(0.2, 7)	1	1
2	MRS	0.79(0.49, 0.96)	0.9(0.69, 0.99)	116.52(5.21, 595.46)	2.57(0.2, 7)	0.93(0.56, 1.32)	1.04(0.78, 1.4)
3	PDFF	0.8(0.69, 0.87)	0.87(0.79, 0.92)	28.88(11.98, 56.75)	1.11(0.2, 5)	0.94(0.75, 1.29)	1.01(0.85, 1.34)
4	TE	0.84(0.73, 0.91)	0.64(0.48, 0.77)	10.7(3.72, 22.69)	0.43(0.14, 1)	0.98(0.82, 1.32)	0.74(0.54, 1)

Notes: Reference citations included in this table could be found in the [supplementary table 6](#).

Abbreviations: HRI, Hepatic Radiation Index; MRS, Magnetic Resonance Spectroscopy; PDFF, Proton Density Fat Fraction; TE, Transient Elastography.

Table 4 40 NITs Diagnostic for NAFL Vs NASH ANOVA Model Analysis Results

No.	NITs	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	S index (95% CI)	Relative Sensitivity (95% CI)	Relative Specificity (95% CI)
1	MRE	0.89(0.68, 0.99)	0.99(0.97, 1)	4,360,974.03(210.74, 3,589,007.06)	63.06(17, 79)	1.01(0.74, 1.45)	1.54(1.03, 3.34)
2	HsCRP	0.63(0.35, 0.84)	0.96(0.77, 1)	942,765.03(4.88, 154,229.3)	10.26(0.2, 47)	0.71(0.38, 1.1)	1.5(0.93, 3.21)
3	IL-1RA	0.84(0.5, 0.98)	0.93(0.62, 1)	1234.67(10.55, 7459.82)	29.77(0.68, 75)	0.95(0.55, 1.41)	1.44(0.79, 3.07)
4	mi-RNA 99a	0.89(0.7, 0.98)	0.93(0.77, 0.99)	282.72(22.63, 1264.09)	33.65(3.73, 73)	1.01(0.75, 1.43)	1.43(0.91, 3.12)
5	AIAT	0.77(0.41, 0.96)	0.76(0.28, 0.99)	110.2(1.27, 513.69)	7.94(0.02, 59)	0.87(0.45, 1.29)	1.19(0.38, 2.81)
6	PC	0.9(0.63, 0.99)	0.72(0.3, 0.96)	109.72(3.48, 668.96)	17.25(0.13, 65)	1	1
7	AC	0.88(0.55, 1)	0.48(0.11, 0.85)	82.36(0.75, 326.64)	5.47(0.02, 41)	1(0.59, 1.47)	0.75(0.15, 1.87)
8	FNI	0.77(0.41, 0.97)	0.82(0.46, 0.98)	55.29(2.46, 292.94)	9.02(0.06, 59)	0.87(0.45, 1.31)	1.28(0.59, 2.87)
9	SAN	0.55(0.19, 0.86)	0.9(0.59, 1)	47.81(1.5, 289.4)	3.11(0.03, 17)	0.62(0.21, 1.07)	1.39(0.76, 3.01)
10	NI-NASH-DS	0.8(0.55, 0.94)	0.83(0.59, 0.97)	42.95(4.08, 182.45)	8.14(0.24, 55)	0.9(0.6, 1.32)	1.29(0.71, 2.82)
11	C13	0.74(0.38, 0.95)	0.78(0.38, 0.97)	28.55(1.63, 139.16)	4.65(0.02, 29)	0.84(0.42, 1.29)	1.2(0.51, 2.75)
12	MR-MASH	0.83(0.52, 0.97)	0.68(0.27, 0.94)	27.46(1.77, 125.68)	6.21(0.03, 47)	0.94(0.56, 1.36)	1.06(0.35, 2.48)
13	mi-RNA 34a	0.76(0.5, 0.93)	0.82(0.57, 0.94)	25.56(3.43, 92.92)	4.43(0.13, 17.67)	0.86(0.55, 1.26)	1.27(0.72, 2.81)
14	ActiTest	0.49(0.15, 0.86)	0.86(0.52, 0.99)	17.68(0.85, 90.34)	1.6(0.02, 9.81)	0.56(0.16, 1.05)	1.33(0.66, 2.91)
15	C-DAG	0.67(0.29, 0.92)	0.76(0.36, 0.97)	16.35(1.1, 74.44)	2.33(0.02, 13)	0.76(0.33, 1.2)	1.18(0.49, 2.67)
16	RBP4	0.72(0.35, 0.95)	0.68(0.26, 0.95)	16.1(0.9, 83.78)	2.77(0.02, 18.33)	0.82(0.39, 1.24)	1.06(0.34, 2.39)
17	FGF21	0.82(0.56, 0.96)	0.63(0.24, 0.88)	14.22(1.25, 55.02)	2.95(0.03, 16.33)	0.92(0.6, 1.35)	0.97(0.34, 2.22)
18	ELF	0.68(0.29, 0.94)	0.68(0.24, 0.95)	12.49(0.58, 62.85)	1.98(0.02, 12.61)	0.77(0.33, 1.21)	1.05(0.31, 2.47)
19	CK-18	0.77(0.57, 0.91)	0.72(0.47, 0.9)	12.2(2.84, 38.19)	2.16(0.07, 9.4)	0.87(0.61, 1.28)	1.12(0.59, 2.46)
20	LEARN	0.84(0.52, 0.97)	0.5(0.13, 0.88)	11.98(0.83, 56.23)	2.37(0.02, 17.67)	0.94(0.57, 1.36)	0.79(0.17, 2)
21	M2BPGi	0.84(0.63, 0.96)	0.54(0.26, 0.82)	11.94(1.43, 48.03)	2.33(0.02, 17)	0.95(0.68, 1.38)	0.85(0.34, 1.99)
22	GGT	0.75(0.48, 0.93)	0.69(0.33, 0.87)	11.24(1.29, 40.64)	1.98(0.02, 11)	0.85(0.53, 1.26)	1.07(0.45, 2.37)
23	NashTest	0.39(0.1, 0.76)	0.86(0.5, 0.99)	10.97(0.54, 56.99)	0.66(0.02, 4.6)	0.44(0.11, 0.92)	1.33(0.65, 2.87)
24	mi-RNA 122	0.72(0.46, 0.91)	0.71(0.43, 0.89)	10.64(1.56, 37.81)	1.49(0.02, 7.29)	0.82(0.5, 1.22)	1.11(0.55, 2.51)
25	AST	0.75(0.48, 0.93)	0.68(0.33, 0.86)	10.53(1.26, 39.36)	1.72(0.02, 9.67)	0.85(0.51, 1.27)	1.05(0.44, 2.36)
26	SWVS	0.58(0.2, 0.9)	0.74(0.33, 0.96)	10.36(0.53, 52.07)	1.31(0.02, 8.21)	0.65(0.22, 1.09)	1.15(0.42, 2.6)
27	TE	0.81(0.69, 0.9)	0.66(0.49, 0.8)	9.33(3.64, 19.94)	1.37(0.16, 5)	0.91(0.73, 1.29)	1.02(0.59, 2.25)
28	FAST	0.8(0.59, 0.92)	0.61(0.36, 0.81)	8.58(1.62, 25.89)	1.26(0.02, 6.43)	0.9(0.63, 1.31)	0.95(0.47, 2.17)
29	ALT	0.66(0.45, 0.83)	0.76(0.5, 0.9)	8.25(1.8, 22.17)	0.91(0.02, 3.86)	0.75(0.48, 1.11)	1.17(0.62, 2.58)
30	CK18-M65	0.75(0.55, 0.9)	0.65(0.41, 0.85)	7.79(1.68, 24.12)	1.01(0.02, 5.29)	0.84(0.59, 1.25)	1.01(0.5, 2.21)
31	AFABP	0.77(0.5, 0.94)	0.56(0.19, 0.83)	7.34(0.64, 26.85)	0.98(0.02, 5.86)	0.87(0.54, 1.27)	0.87(0.25, 2)
32	BARD	0.73(0.45, 0.91)	0.62(0.33, 0.85)	7.32(1.01, 27.38)	1.01(0.02, 6.6)	0.83(0.49, 1.25)	0.97(0.43, 2.22)
33	TyG	0.67(0.28, 0.93)	0.59(0.18, 0.91)	6.45(0.44, 29.33)	0.94(0.01, 6.43)	0.76(0.31, 1.21)	0.92(0.23, 2.22)
34	ION	0.68(0.3, 0.93)	0.58(0.18, 0.9)	6.02(0.45, 27.65)	0.83(0.01, 5.86)	0.77(0.34, 1.22)	0.9(0.24, 2.23)
35	CK18-M30	0.58(0.42, 0.73)	0.78(0.63, 0.89)	5.72(2.11, 13.06)	0.4(0.03, 1.67)	0.66(0.45, 0.99)	1.21(0.73, 2.62)
36	APRI	0.64(0.43, 0.8)	0.72(0.52, 0.87)	5.66(1.52, 14.08)	0.49(0.02, 2.11)	0.72(0.47, 1.07)	1.12(0.63, 2.53)
37	FIB4	0.72(0.57, 0.84)	0.61(0.44, 0.77)	4.83(1.82, 10.46)	0.34(0.02, 1.4)	0.82(0.61, 1.2)	0.96(0.54, 2.15)
38	HAALT scoring system	0.62(0.34, 0.88)	0.62(0.4, 0.85)	4.21(0.65, 16.14)	0.33(0.02, 2.34)	0.7(0.38, 1.11)	0.97(0.49, 2.18)
39	GP73	0.92(0.77, 0.98)	0.2(0.06, 0.48)	4.15(0.59, 14.43)	0.6(0.03, 3.57)	1.04(0.81, 1.49)	0.32(0.08, 0.87)
40	NFS	0.62(0.44, 0.77)	0.68(0.49, 0.83)	4.09(1.3, 9.43)	0.25(0.02, 1.11)	0.7(0.47, 1.05)	1.06(0.6, 2.37)

Notes: Reference citations included in this table could be found in the [supplementary eTable 7](#).

Abbreviation: List of abbreviations were provided in [Supplemental eTable 1](#).

Quality of Evidence

Two authors critically appraised the 180 studies using the critical criteria of the QUADAS-2 assessment criteria. Details are shown in [Figure 7](#). For applicability concerns, the included studies aligned well with the review questions. Specially, 100% (180/180) of reference standard, 97.2% (175/180) of patient selection, 97.8% (176/180) of index test had low applicability concerns. However,

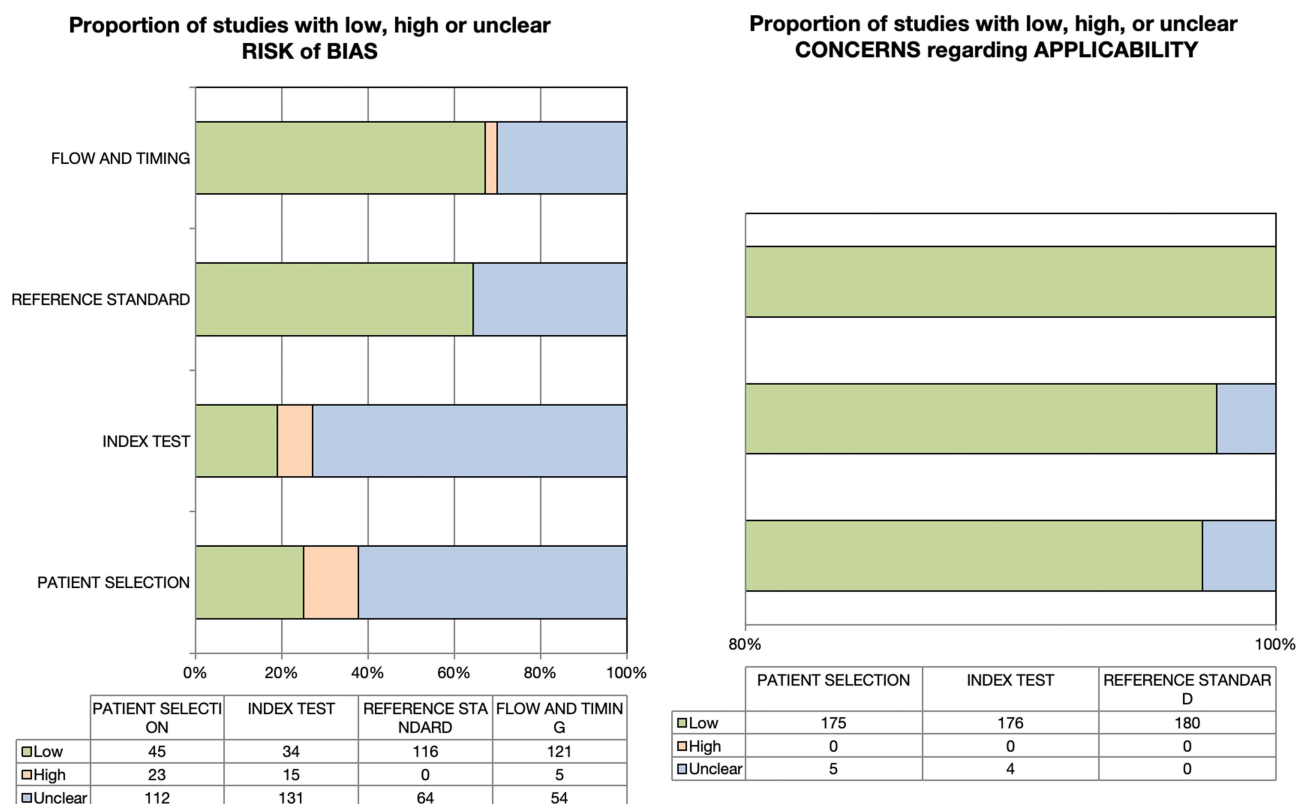


Figure 7 QUADAS-2 quality evaluation results of 180 included studies.

regarding the risk of bias, 75.0% (135/180) of the studies needed a clear description (unclear or high risk), with only 23.3% (42/180) being properly reported. The same issue occurred in the index test domain, where 81.1% (146/180) of the studies needed more detailed descriptions of NITs.

Table 5 15 NITs Diagnostic for NAFLD With Fibrosis ANOVA Model Analysis Results

No.	NITs	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	S index (95% CI)	Relative Sensitivity (95% CI)	Relative Specificity (95% CI)
1	sH2a	0.72(0.32, 0.95)	0.72(0.25, 0.98)	40.64(0.64, 244.34)	9.01(0.08, 27)	1.17(0.43, 2.88)	1.38(0.36, 4.09)
2	IgA	0.73(0.43, 0.95)	0.73(0.26, 0.97)	27.38(0.78, 157.05)	8.92(0.08, 27)	1.18(0.56, 2.88)	1.39(0.39, 4.13)
3	siCAM-I	0.71(0.31, 0.94)	0.67(0.2, 0.97)	23.18(0.45, 152.53)	6.41(0.05, 27)	1.15(0.38, 2.8)	1.29(0.29, 3.83)
4	AAR score	0.7(0.28, 0.96)	0.63(0.19, 0.96)	18.24(0.33, 108.01)	5.95(0.05, 25)	1	1
5	NIS	0.63(0.31, 0.86)	0.66(0.22, 0.96)	10.13(0.34, 59.72)	3.99(0.05, 23)	1.01(0.42, 2.34)	1.26(0.32, 3.75)
6	TE	0.6(0.45, 0.77)	0.81(0.55, 0.95)	9.98(1.7, 32.59)	4.28(0.2, 19)	0.98(0.55, 2.31)	1.55(0.73, 4.34)
7	ASMase	0.74(0.53, 0.91)	0.67(0.33, 0.91)	9.93(1.14, 36.86)	5.79(0.09, 23)	1.2(0.64, 2.85)	1.27(0.45, 3.73)
8	Hui score	0.6(0.33, 0.85)	0.69(0.25, 0.97)	9.37(0.38, 51.83)	3.26(0.05, 21)	0.97(0.43, 2.29)	1.32(0.38, 3.9)
9	FIB4	0.68(0.55, 0.81)	0.78(0.6, 0.88)	9.16(2.81, 21.7)	5.13(0.43, 19)	1.1(0.63, 2.64)	1.5(0.75, 4.12)
10	Forns index	0.63(0.36, 0.87)	0.61(0.17, 0.94)	6.35(0.3, 33.96)	2.67(0.05, 19)	1.02(0.47, 2.41)	1.16(0.26, 3.5)
11	APRI	0.54(0.37, 0.67)	0.73(0.53, 0.87)	3.89(1.15, 9.48)	1.16(0.08, 5.67)	0.86(0.48, 1.84)	1.39(0.7, 3.82)
12	BARD	0.84(0.54, 0.95)	0.29(0.08, 0.65)	3.85(0.28, 16.3)	1.89(0.06, 11)	1.34(0.73, 2.89)	0.55(0.12, 1.69)
13	Fasting C-peptide	0.64(0.34, 0.9)	0.42(0.07, 0.83)	2.9(0.1, 14.45)	1.14(0.04, 11)	1.04(0.43, 2.57)	0.82(0.1, 2.82)
14	BAAT	0.04(0, 0.28)	0.73(0.23, 0.99)	0.44(0, 2.89)	0.44(0.04, 3)	0.06(0, 0.49)	1.39(0.36, 4.06)
15	NFS	0.04(0, 0.13)	0.58(0.25, 0.86)	0.09(0, 0.42)	0.12(0.04, 0.6)	0.07(0.01, 0.26)	1.11(0.36, 3.29)

Notes: Reference citations included in this table could be found in the [supplementary table 8](#).

Abbreviation: List of abbreviations were provided in [Supplemental eTable 1](#).

Table 6 40 NITs Diagnostic for NAFLD ≥F2 ANOVA Model Analysis Results

No.	NITs	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	S index (95% CI)	Relative Sensitivity (95% CI)	Relative Specificity (95% CI)
1	HRI	0.84(0.55, 0.98)	0.83(0.51, 0.98)	80.94(6.46, 391.41)	39.24(0.71, 77)	1.04(0.68, 1.24)	1(0.61, 1.22)
2	HSI	0.6(0.37, 0.8)	0.97(0.92, 0.99)	63.75(13.13, 196.85)	21.16(1, 61)	0.74(0.46, 1)	1.17(1.08, 1.26)
3	SHG B-index	0.8(0.5, 0.96)	0.82(0.51, 0.97)	40.57(4.84, 160.84)	27.2(0.33, 73)	0.98(0.61, 1.21)	0.99(0.61, 1.21)
4	FLI	0.77(0.6, 0.89)	0.86(0.7, 0.96)	30.75(6.37, 93.26)	26.37(0.76, 69)	0.94(0.75, 1.11)	1.04(0.85, 1.18)
5	Ali2021	0.81(0.54, 0.96)	0.79(0.48, 0.95)	28.01(4.85, 89.58)	21.35(0.33, 69)	0.99(0.66, 1.2)	0.95(0.58, 1.18)
6	LRM	0.81(0.65, 0.91)	0.83(0.7, 0.91)	25.45(6.99, 64.38)	23.81(1.13, 69)	0.99(0.8, 1.14)	1(0.83, 1.13)
7	MRE	0.81(0.76, 0.86)	0.83(0.77, 0.88)	22.36(13.17, 36.15)	20.61(3.89, 61)	1	1
8	2D-SWE	0.91(0.82, 0.97)	0.61(0.38, 0.81)	21.72(5.59, 61.32)	13.41(0.71, 53)	1.12(0.99, 1.23)	0.74(0.47, 1)
9	Fibro-3	0.81(0.62, 0.93)	0.77(0.57, 0.9)	19.45(4.79, 54.81)	15.02(0.33, 65)	0.99(0.76, 1.17)	0.93(0.67, 1.12)
10	ElastPQ	0.68(0.46, 0.85)	0.83(0.64, 0.94)	14.13(2.95, 41.85)	8.44(0.07, 53)	0.84(0.57, 1.06)	1(0.78, 1.16)
11	NIS	0.8(0.67, 0.89)	0.73(0.58, 0.86)	13.55(4.35, 32.69)	7.83(0.26, 51)	0.98(0.82, 1.12)	0.89(0.7, 1.06)
12	FIB8	0.86(0.72, 0.95)	0.58(0.39, 0.77)	11.41(2.71, 33.59)	5.05(0.1, 33)	1.06(0.88, 1.2)	0.7(0.47, 0.94)
13	PDFF	0.86(0.58, 0.98)	0.44(0.14, 0.78)	9.9(1.05, 41.87)	3.63(0.02, 19)	1.06(0.72, 1.25)	0.53(0.17, 0.95)
14	ARFI	0.67(0.54, 0.77)	0.81(0.72, 0.89)	9.28(4.83, 15.81)	2.8(0.24, 11.67)	0.82(0.66, 0.97)	0.98(0.85, 1.11)
15	Hepascore	0.6(0.41, 0.77)	0.82(0.67, 0.92)	8.76(2.4, 23.65)	2.88(0.03, 19)	0.74(0.5, 0.95)	0.99(0.81, 1.14)
16	CA index	0.63(0.44, 0.79)	0.8(0.65, 0.91)	8.67(2.37, 22.95)	3.05(0.03, 17.67)	0.77(0.54, 0.99)	0.97(0.79, 1.12)
17	TE	0.72(0.67, 0.76)	0.76(0.71, 0.8)	8.29(6.01, 11.18)	1.84(0.53, 4.43)	0.88(0.81, 0.96)	0.92(0.85, 1)
18	ELF	0.68(0.57, 0.78)	0.78(0.68, 0.86)	8.19(3.99, 15.21)	2.18(0.15, 9)	0.84(0.7, 0.98)	0.94(0.81, 1.07)
19	PROC3	0.62(0.51, 0.71)	0.8(0.72, 0.87)	7.04(3.92, 11.69)	1.29(0.11, 5)	0.76(0.62, 0.89)	0.97(0.85, 1.09)
20	P4NP 7S	0.7(0.51, 0.85)	0.7(0.53, 0.84)	6.85(1.92, 18.34)	1.9(0.02, 11.67)	0.86(0.63, 1.05)	0.84(0.63, 1.03)
21	FIB4	0.67(0.61, 0.72)	0.77(0.72, 0.81)	6.73(4.88, 9.17)	1.05(0.29, 2.56)	0.82(0.74, 0.9)	0.93(0.85, 1.01)
22	Hepamet	0.74(0.57, 0.87)	0.65(0.47, 0.8)	6.53(1.97, 16.97)	1.6(0.02, 9.67)	0.91(0.69, 1.09)	0.78(0.57, 0.98)
23	M2BPGi	0.73(0.56, 0.88)	0.64(0.46, 0.8)	6.24(1.75, 16.84)	1.44(0.02, 8.6)	0.9(0.67, 1.1)	0.77(0.55, 0.97)
24	Marta2020	0.64(0.31, 0.89)	0.68(0.34, 0.92)	5.97(0.99, 19.31)	1.6(0.02, 11)	0.78(0.38, 1.11)	0.83(0.41, 1.13)
25	FibroTest	0.7(0.55, 0.83)	0.69(0.54, 0.82)	5.94(2.3, 13.05)	1.05(0.03, 5)	0.86(0.67, 1.04)	0.83(0.64, 1)
26	RPR	0.49(0.17, 0.81)	0.79(0.45, 0.96)	5.9(0.86, 20.42)	1.2(0.02, 8.33)	0.61(0.21, 1.01)	0.95(0.54, 1.19)
27	CTL-S	0.66(0.49, 0.8)	0.71(0.54, 0.84)	5.64(1.78, 13.69)	1.02(0.02, 5.45)	0.82(0.6, 1)	0.85(0.65, 1.03)
28	NFS	0.64(0.56, 0.71)	0.75(0.68, 0.81)	5.5(3.61, 8.14)	0.58(0.1, 1.77)	0.79(0.69, 0.89)	0.9(0.81, 1)
29	APRI	0.7(0.63, 0.77)	0.68(0.6, 0.74)	5.18(3.3, 7.74)	0.52(0.1, 1.53)	0.87(0.76, 0.97)	0.82(0.72, 0.92)
30	p-SWE	0.61(0.35, 0.83)	0.7(0.48, 0.87)	5.1(0.99, 15.46)	1.19(0.02, 7.67)	0.75(0.43, 1.03)	0.85(0.58, 1.07)
31	MASEF	0.72(0.51, 0.87)	0.6(0.38, 0.8)	4.92(1.36, 12.71)	0.83(0.02, 4.72)	0.88(0.62, 1.09)	0.73(0.46, 0.98)
32	Forns index	0.63(0.42, 0.81)	0.69(0.51, 0.83)	4.61(1.25, 12.84)	0.68(0.02, 4.11)	0.77(0.51, 1)	0.83(0.61, 1.02)
33	AAR score	0.55(0.37, 0.73)	0.75(0.58, 0.88)	4.6(1.31, 11.84)	0.59(0.02, 3.67)	0.67(0.45, 0.9)	0.91(0.71, 1.07)
34	MEFIB	0.36(0.17, 0.59)	0.85(0.68, 0.95)	4.44(0.9, 13.62)	0.37(0.02, 2.27)	0.45(0.21, 0.73)	1.03(0.81, 1.19)
35	FAST	0.6(0.45, 0.73)	0.73(0.6, 0.83)	4.35(2.05, 8.5)	0.37(0.02, 1.73)	0.74(0.55, 0.91)	0.88(0.73, 1.02)
36	Clinical-CT index	0.57(0.38, 0.74)	0.72(0.55, 0.85)	4.09(1.25, 10.11)	0.42(0.02, 2.56)	0.7(0.48, 0.91)	0.87(0.67, 1.04)
37	BARD	0.6(0.47, 0.73)	0.62(0.48, 0.74)	2.68(1.13, 5.52)	0.1(0.02, 0.45)	0.74(0.57, 0.91)	0.75(0.58, 0.91)
38	FIB6	0.82(0.55, 0.96)	0.22(0.05, 0.51)	1.67(0.35, 5.06)	0.49(0.02, 3)	1.01(0.67, 1.21)	0.27(0.06, 0.63)
39	ALT	0.65(0.32, 0.91)	0.28(0.07, 0.61)	0.93(0.16, 3.18)	0.11(0.01, 0.71)	0.8(0.39, 1.13)	0.34(0.09, 0.74)
40	M2BP	0.8(0.63, 0.91)	0.05(0.02, 0.12)	0.27(0.07, 0.76)	0.08(0.02, 0.33)	0.99(0.78, 1.14)	0.07(0.03, 0.15)

Notes: Reference citations included in this table could be found in the [supplementary eTable 9](#).

Abbreviation: List of abbreviations were provided in [Supplemental eTable 1](#).

Discussion

To serve as a reference for selecting non-invasive diagnostic tools for identifying different stages of NAFLD in clinical practice, we compiled and analyzed diagnostic performance data from various non-invasive tests found in the existing literature and ranked their accuracy. In the context of diagnosing steatosis, our analysis revealed that for S1 steatosis, H-MRS demonstrated the highest diagnostic efficacy, followed by NLV, PDFF, HSI, VTQ, MRI, UPGAP, and UAP, with TE showing the lowest diagnostic efficacy. Among serum markers, ALT, TyG, FAI, VAI, and CK18-M30 showed higher diagnostic accuracy. For S2 steatosis, HRI exhibited the highest diagnostic accuracy, followed by MRS, PDFF, and TE.

Table 7 64 NITs Diagnostic for Advanced Fibrosis ($\geq F3$) ANOVA Model Analysis Results

No.	NITs	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	S index (95% CI)	Relative Sensitivity (95% CI)	Relative Specificity (95% CI)
1	CK-18	0.95(0.65, 1)	0.53(0.19, 0.85)	102.654.16(1.6, 134,059.8)	21.95(0.1, 97)	1.11(0.75, 1.29)	0.61(0.21, 0.98)
2	SHG B-index	0.81(0.44, 0.98)	0.95(0.77, 1)	1149.9(11.35, 5983.64)	73.13(2.43, 125)	0.94(0.51, 1.19)	1.09(0.89, 1.24)
3	LFS	0.83(0.41, 0.99)	0.89(0.58, 1)	793.48(4.17, 5316.76)	57.77(0.45, 125)	0.97(0.47, 1.23)	1.02(0.66, 1.21)
4	LPR	0.89(0.61, 1)	0.78(0.49, 0.95)	239.87(3.96, 1371.41)	40.27(0.41, 117)	1.04(0.7, 1.24)	0.9(0.56, 1.12)
5	FM	0.87(0.66, 0.98)	0.8(0.5, 0.96)	78.67(4.6, 382.09)	34.26(0.58, 117)	1.02(0.76, 1.21)	0.92(0.57, 1.13)
6	HSI	0.72(0.39, 0.94)	0.89(0.66, 0.98)	61.12(3.53, 306.22)	26.31(0.33, 113)	0.84(0.45, 1.13)	1.02(0.75, 1.19)
7	FLI	0.83(0.57, 0.97)	0.79(0.47, 0.96)	52.18(3.43, 247.93)	24.61(0.3, 113)	0.97(0.67, 1.18)	0.91(0.55, 1.13)
8	MRE	0.86(0.76, 0.93)	0.88(0.79, 0.94)	51.11(18.71, 115.19)	31.23(5.15, 105)	1(1, 1)	1(1, 1)
9	10 metabolites	0.73(0.33, 0.97)	0.84(0.59, 0.97)	50.89(1.88, 271.77)	20.12(0.12, 111)	0.85(0.39, 1.17)	0.96(0.65, 1.16)
10	Hepascore	0.82(0.67, 0.92)	0.88(0.75, 0.95)	47.17(10.81, 141.89)	26.76(2.52, 107)	0.96(0.77, 1.13)	1.01(0.84, 1.15)
11	PLT	0.82(0.49, 0.98)	0.64(0.32, 0.89)	28.86(1.16, 159.52)	9.68(0.06, 85.05)	0.96(0.56, 1.2)	0.74(0.37, 1.05)
12	HA	0.87(0.62, 0.98)	0.6(0.27, 0.88)	28.59(1.61, 141.33)	10.11(0.09, 77)	1.02(0.7, 1.22)	0.69(0.31, 1.02)
13	NFI	0.8(0.51, 0.96)	0.68(0.34, 0.92)	20.92(1.26, 96.99)	8.71(0.06, 81)	0.94(0.59, 1.17)	0.78(0.38, 1.07)
14	RPR	0.7(0.3, 0.95)	0.75(0.41, 0.95)	20.45(1.04, 100.87)	8.64(0.04, 87)	0.82(0.35, 1.14)	0.86(0.46, 1.13)
15	P4NP 7S	0.72(0.53, 0.87)	0.85(0.71, 0.94)	19.85(4.65, 56.55)	8.53(0.55, 35)	0.84(0.6, 1.06)	0.97(0.8, 1.12)
16	NAFLD-FM	0.66(0.35, 0.9)	0.81(0.53, 0.96)	19.52(1.53, 86.03)	8.53(0.08, 81.05)	0.78(0.41, 1.07)	0.93(0.6, 1.14)
17	Agile 3+	0.79(0.58, 0.93)	0.77(0.56, 0.92)	19.47(3.37, 64.51)	8.44(0.3, 69)	0.92(0.67, 1.13)	0.88(0.63, 1.09)
18	S Index	0.7(0.41, 0.91)	0.78(0.49, 0.95)	18.47(1.53, 82.15)	7.99(0.09, 79.05)	0.82(0.48, 1.11)	0.89(0.55, 1.12)
19	ElastPQ	0.77(0.45, 0.95)	0.69(0.35, 0.92)	17.93(1.18, 81.86)	7.54(0.06, 79)	0.9(0.52, 1.16)	0.79(0.4, 1.08)
20	TE	0.84(0.79, 0.87)	0.77(0.71, 0.82)	17.58(11.48, 25.71)	5.77(2.05, 14.6)	0.98(0.89, 1.11)	0.88(0.79, 0.99)
21	ARFI	0.68(0.47, 0.85)	0.81(0.64, 0.93)	12.79(2.93, 36.44)	4.25(0.23, 19)	0.8(0.54, 1.02)	0.93(0.73, 1.11)
22	ADAPT	0.73(0.41, 0.93)	0.68(0.35, 0.92)	12.56(0.92, 60)	4.69(0.04, 30.33)	0.86(0.48, 1.12)	0.78(0.39, 1.07)
23	Autotaxin	0.48(0.17, 0.8)	0.86(0.61, 0.97)	12.53(0.89, 55.27)	4.48(0.03, 31)	0.57(0.2, 0.94)	0.98(0.69, 1.16)
24	PRO-C6	0.72(0.4, 0.92)	0.69(0.37, 0.92)	12.42(1, 55.72)	4.67(0.05, 30.33)	0.85(0.46, 1.11)	0.79(0.42, 1.07)
25	PRO-C4	0.41(0.13, 0.74)	0.87(0.63, 0.98)	11.29(0.73, 52.16)	3.48(0.02, 23)	0.48(0.15, 0.87)	1(0.72, 1.18)
26	FIBC-3	0.6(0.26, 0.87)	0.77(0.47, 0.96)	11.2(0.79, 54.02)	4.1(0.03, 28.33)	0.71(0.29, 1.03)	0.88(0.52, 1.12)
27	MACK-3	0.7(0.37, 0.91)	0.69(0.35, 0.93)	11.08(0.78, 50.83)	3.99(0.03, 27)	0.82(0.43, 1.1)	0.79(0.4, 1.08)
28	SMPA	0.7(0.36, 0.93)	0.67(0.33, 0.92)	11(0.78, 51.3)	3.9(0.03, 27.03)	0.82(0.43, 1.12)	0.77(0.37, 1.07)
29	Caussy 2018	0.73(0.34, 0.97)	0.56(0.22, 0.86)	10.2(0.45, 51.52)	2.97(0.01, 23.67)	0.86(0.4, 1.17)	0.64(0.25, 0.99)
30	CAO2013	0.55(0.22, 0.83)	0.79(0.5, 0.96)	10.04(0.75, 48.71)	3.39(0.03, 23.67)	0.64(0.26, 1)	0.91(0.56, 1.14)
31	M2BPGi	0.82(0.54, 0.96)	0.5(0.19, 0.81)	9.42(0.72, 44.14)	2.91(0.03, 23)	0.95(0.62, 1.18)	0.58(0.22, 0.94)
32	FibroTest	0.67(0.44, 0.84)	0.76(0.56, 0.91)	9.34(1.78, 28.39)	2.64(0.11, 12.6)	0.78(0.51, 1.01)	0.88(0.63, 1.07)
33	CK18-M30	0.57(0.24, 0.85)	0.76(0.44, 0.95)	9.29(0.65, 39.99)	2.91(0.02, 19.02)	0.67(0.29, 1.01)	0.87(0.5, 1.11)
34	FAST	0.57(0.25, 0.85)	0.77(0.47, 0.95)	9.2(0.72, 40.25)	3.04(0.02, 21.67)	0.66(0.28, 1.01)	0.89(0.53, 1.11)
35	SWE	0.73(0.6, 0.84)	0.73(0.61, 0.84)	8.65(3.51, 17.91)	1.88(0.31, 6.11)	0.86(0.69, 1.03)	0.84(0.68, 1)
36	BAAT	0.88(0.64, 0.99)	0.31(0.08, 0.64)	8.29(0.44, 42.61)	1.72(0.02, 13)	1.03(0.75, 1.23)	0.35(0.09, 0.73)
37	FibroMeter NAFLD Score	0.65(0.41, 0.84)	0.75(0.52, 0.91)	8.11(1.47, 26.98)	2.2(0.07, 11.29)	0.76(0.47, 1)	0.86(0.58, 1.07)
38	THE	0.79(0.48, 0.96)	0.48(0.17, 0.81)	8(0.46, 37.92)	2.33(0.02, 15.67)	0.93(0.56, 1.17)	0.55(0.19, 0.93)
39	ELF	0.77(0.58, 0.9)	0.65(0.44, 0.82)	7.84(1.98, 20.72)	1.64(0.12, 7.29)	0.9(0.67, 1.1)	0.74(0.5, 0.96)
40	Fibrometer V2G	0.68(0.45, 0.86)	0.7(0.45, 0.88)	7.08(1.35, 22.56)	1.57(0.06, 8.11)	0.8(0.52, 1.03)	0.8(0.51, 1.03)
41	IgA	0.7(0.36, 0.93)	0.59(0.25, 0.87)	6.9(0.53, 30.58)	2.04(0.01, 12.72)	0.82(0.42, 1.11)	0.67(0.29, 1.01)
42	NAFLD score	0.62(0.27, 0.9)	0.66(0.32, 0.92)	6.89(0.47, 31.02)	2.14(0.01, 13.81)	0.73(0.32, 1.07)	0.76(0.36, 1.06)
43	aMAP	0.41(0.13, 0.74)	0.83(0.58, 0.97)	6.81(0.46, 29.59)	1.84(0.01, 12.61)	0.48(0.15, 0.88)	0.95(0.66, 1.16)
44	BAAAP	0.52(0.2, 0.83)	0.75(0.42, 0.94)	6.42(0.49, 27.45)	1.88(0.01, 12.34)	0.61(0.24, 0.99)	0.86(0.47, 1.1)
45	Hui score	0.4(0.14, 0.72)	0.82(0.55, 0.97)	6.19(0.49, 26.8)	1.6(0.01, 10.34)	0.47(0.16, 0.86)	0.94(0.62, 1.14)
46	PRO-C3	0.68(0.55, 0.8)	0.72(0.6, 0.82)	6.09(2.81, 11.94)	1.03(0.19, 3.13)	0.8(0.63, 0.97)	0.82(0.67, 0.97)
47	AP index	0.49(0.2, 0.79)	0.77(0.49, 0.95)	6.06(0.48, 25.97)	1.57(0.01, 11)	0.57(0.23, 0.93)	0.88(0.55, 1.11)
48	FIB4	0.61(0.55, 0.67)	0.79(0.75, 0.83)	6.04(4.31, 8.17)	1.02(0.4, 2.16)	0.72(0.62, 0.83)	0.9(0.82, 1.01)
49	NFS	0.64(0.58, 0.7)	0.76(0.71, 0.81)	5.83(4.02, 8.23)	0.95(0.35, 2.05)	0.75(0.65, 0.87)	0.87(0.79, 0.98)
50	N score	0.62(0.3, 0.87)	0.62(0.31, 0.88)	4.87(0.42, 22.13)	1.1(0.01, 6.82)	0.72(0.35, 1.05)	0.71(0.35, 1.03)
51	NIKEI	0.44(0.15, 0.78)	0.75(0.43, 0.94)	4.86(0.34, 23.68)	1.23(0.01, 8.72)	0.52(0.18, 0.91)	0.86(0.49, 1.11)

(Continued)

Table 7 (Continued).

No.	NITs	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	S index (95% CI)	Relative Sensitivity (95% CI)	Relative Specificity (95% CI)
52	APRI	0.58(0.51, 0.65)	0.77(0.72, 0.81)	4.71(3.23, 6.69)	0.63(0.22, 1.4)	0.68(0.58, 0.8)	0.88(0.79, 0.99)
53	PDFF	0.6(0.21, 0.93)	0.57(0.21, 0.88)	4.63(0.24, 22.46)	1.29(0.01, 8.33)	0.7(0.24, 1.1)	0.65(0.23, 1.03)
54	NAFIC	0.51(0.19, 0.83)	0.69(0.37, 0.91)	4.45(0.33, 20.57)	1.02(0.01, 6.6)	0.59(0.23, 0.98)	0.79(0.42, 1.07)
55	Prothrombin-INR	0.65(0.26, 0.94)	0.48(0.16, 0.81)	4.37(0.19, 22.56)	1.05(0.01, 7.67)	0.77(0.3, 1.13)	0.55(0.19, 0.93)
56	BARD	0.65(0.57, 0.73)	0.66(0.59, 0.73)	3.86(2.32, 5.94)	0.43(0.13, 1)	0.77(0.65, 0.9)	0.76(0.65, 0.87)
57	FIB6	0.79(0.46, 0.96)	0.31(0.07, 0.67)	3.26(0.22, 15.46)	0.61(0.01, 3.67)	0.92(0.54, 1.17)	0.35(0.08, 0.78)
58	Albumin	0.65(0.25, 0.95)	0.37(0.1, 0.71)	2.85(0.11, 17.2)	0.62(0.01, 4.07)	0.76(0.29, 1.13)	0.43(0.12, 0.82)
59	AAR score	0.54(0.45, 0.64)	0.69(0.6, 0.77)	2.75(1.52, 4.57)	0.24(0.06, 0.62)	0.64(0.51, 0.78)	0.79(0.67, 0.92)
60	CK18-M65	0.41(0.13, 0.74)	0.68(0.36, 0.91)	2.64(0.21, 11.79)	0.38(0.01, 2.64)	0.48(0.15, 0.86)	0.78(0.4, 1.06)
61	GUCI index	0.54(0.27, 0.79)	0.59(0.34, 0.82)	2.41(0.37, 8.02)	0.28(0.01, 1.53)	0.63(0.31, 0.94)	0.68(0.38, 0.95)
62	KING score	0.35(0.09, 0.68)	0.7(0.36, 0.93)	2.34(0.16, 11.14)	0.36(0.01, 2.58)	0.41(0.1, 0.79)	0.81(0.41, 1.08)
63	Forns index	0.56(0.32, 0.78)	0.44(0.22, 0.7)	1.32(0.24, 4.23)	0.11(0.01, 0.48)	0.66(0.37, 0.93)	0.51(0.25, 0.8)
64	Lok index	0.32(0.08, 0.66)	0.31(0.07, 0.64)	0.3(0.02, 1.33)	0.02(0.01, 0.07)	0.38(0.09, 0.78)	0.35(0.08, 0.73)

Notes: Reference citations included in this table could be found in the [supplementary table 10](#).

Abbreviation: List of abbreviations were provided in [Supplemental eTable 1](#).

Table 8 14 NITs Diagnostic for High-Risk NASH (NAS \geq 4, \geq F2) ANOVA Model Analysis Results

No.	NITs	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	S index (95% CI)	Relative Sensitivity (95% CI)	Relative Specificity (95% CI)
1	CK-18	0.88(0.52, 1)	0.37(0.07, 0.76)	39.46(0.48, 267.09)	2.78(0.05, 17)	1.28(0.64, 2.62)	0.67(0.11, 2.07)
2	ELF	0.99(0.96, 1)	0.07(0.01, 0.2)	34.64(0.83, 181.08)	1.03(0.33, 3)	1.43(1.01, 2.94)	0.13(0.02, 0.51)
3	C-DAG	0.71(0.32, 0.96)	0.8(0.43, 0.97)	29.34(1.59, 144.73)	8.26(0.09, 23)	1.03(0.39, 2.32)	1.46(0.62, 3.67)
4	RTE	0.76(0.34, 0.97)	0.63(0.22, 0.93)	18.1(0.7, 96.33)	5.21(0.05, 21)	1	1
5	FNI	0.66(0.44, 0.82)	0.85(0.61, 0.96)	17.36(2.48, 59.67)	6.53(0.2, 17)	0.96(0.52, 2.04)	1.55(0.78, 3.81)
6	MRE	0.7(0.28, 0.95)	0.66(0.25, 0.93)	11.27(0.64, 53.29)	3.76(0.05, 19)	1.01(0.35, 2.32)	1.21(0.38, 3.28)
7	MACK-3	0.69(0.51, 0.86)	0.75(0.5, 0.93)	10.63(1.82, 38.35)	4.47(0.08, 17)	1(0.58, 2.09)	1.38(0.65, 3.51)
8	MEFIB	0.8(0.62, 0.9)	0.6(0.3, 0.85)	8.34(1.47, 27.15)	3.83(0.09, 17)	1.15(0.71, 2.43)	1.1(0.42, 2.95)
9	NIS	0.79(0.64, 0.9)	0.61(0.32, 0.83)	7.9(1.62, 23.06)	3.64(0.11, 15)	1.14(0.7, 2.39)	1.13(0.45, 2.99)
10	TE	0.86(0.74, 0.94)	0.45(0.25, 0.64)	6.36(1.77, 16.52)	1.8(0.14, 9)	1.25(0.82, 2.59)	0.83(0.33, 2.21)
11	FAST	0.63(0.51, 0.74)	0.72(0.6, 0.81)	4.67(2.27, 8.35)	0.93(0.08, 5)	0.92(0.56, 1.93)	1.32(0.71, 3.3)
12	APRI	0.6(0.45, 0.74)	0.69(0.48, 0.83)	3.86(1.27, 9)	0.52(0.06, 3)	0.87(0.5, 1.89)	1.26(0.62, 3.21)
13	FIB4	0.62(0.48, 0.74)	0.68(0.52, 0.8)	3.78(1.57, 7.51)	0.5(0.07, 2.33)	0.89(0.53, 1.9)	1.25(0.64, 3.13)
14	NFS	0.7(0.56, 0.83)	0.52(0.32, 0.69)	2.93(0.97, 6.55)	0.33(0.06, 1.67)	1.01(0.62, 2.16)	0.97(0.44, 2.55)

Notes: Reference citations included in this table could be found in the [supplementary table 11](#).

Abbreviation: List of abbreviations were provided in [Supplemental eTable 1](#).

However, further evidence is required to validate the diagnosis of S2 steatosis using serological NITs. A meta-analysis by Yokoo et al indicated that MRS had higher accuracy in diagnosing mild hepatic steatosis than other NITs, achieving sensitivity and specificity rates of 77–95% and 81–97%, respectively.²⁷ Nonetheless, the clinical application of MRS is limited by its long scanning time, high cost, and potential for sampling error. In clinical practice, TE is frequently utilized as an NIT for screening hepatic steatosis and fibrosis. However, our study suggests that TE has low sensitivity for detecting early steatosis (S1), which aligns with previous findings.^{28,29} This lower sensitivity may be attributed to its firm reliance on operator skill, inability to measure liver fat content precisely, and susceptibility to obesity, which can diminish diagnostic accuracy. MRE showed the highest efficacy in distinguishing between NAFL and NASH, followed by various serological markers, with HsCRP and IL-1RA ranking amongst the top three. In our study, ION ranked 34th with a DOR of 6.02 (95% CI 0.45–27.65). A previous study suggested that ION could effectively differentiate between MASLD and

MASH, achieving an AUC of nearly 0.85. This best performance may stem from their study's defining of specific features of MASH, such as hepatocellular ballooning and lobular inflammation.³⁰ Additionally, the evidence map (Figure 2) indicates insufficient evidence for diagnosing S2 steatosis using serological NITs. It may be related to the inadequate sensitivity of the serological markers for S2, particularly in the early stages of inflammation before it progresses to overt liver fibrosis. NAFLD is a dynamic disease, and S2 steatosis signifies an increased fat load. Factors initiating inflammation and fibrosis—such as reactive oxygen species and pro-fibrotic agents—may be significantly heightened during this stage. This increase could interfere with serological NITs, making them less accurate in assessing S2 steatosis and more focused on evaluating fibrosis progression. This could help explain the lack of robust evidence, and further mechanistic studies are needed to support this hypothesis.

The largest number of NITs studies diagnosed advanced fibrosis, revealing up to 64 tests identified within the advanced fibrosis subgroup. Among the serological markers, CK18, SHG B-index and LFS had the highest overall performance. Among imaging-based NITs, HSI, MRE and TE ranked the top three. ELF, the first FDA-approved test for the assessing advanced fibrosis, ranked 39th in our NMA (DOR 7.84, 95% CI 1.98–20.72). In addition, ELF is included in the AACE and AGA guidelines for identifying “high-risk NASH”. Our results support this approach, as ELF showed strong performance in the high-risk NASH subgroup, ranked 2th with a DOR of 34.64 (95% CI 0.83 to 181.08), just behind CK18. An analysis showed that FIB-4 was superior to APRI as a predictor of advanced fibrosis, with a validated AUC (FIB-4 vs APRI 0.802 vs 0.73),³¹ which is consistent with our conclusion (FIB-4 vs APRI 0.84 vs 0.79).

In the subgroup with significant fibrosis, the HRI, HSI, and MRE demonstrated high diagnostic accuracy in identifying this condition. The SHG B-index, FLI, Ali2021, and LRM exhibited the highest diagnostic efficacy among the serological markers. While one study indicated that the APRI test was the most effective NIT for identifying at least F2 fibrosis (with an AUC of 0.735),³⁰ we found that APRI identified at least F2 fibrosis with an AUC of 0.75 (95% CI 0.71–0.79). The DOR for indirect comparisons was 5.18 (95% CI 3.30–7.74). It indicates a close value; however, other serological NITs with higher diagnostic accuracy were identified in this study.

In diagnosing high-risk NASH, the three serological markers that demonstrated the best overall performance were CK-18, ELF, C-DAG. Among the imaging-based NITs, transient elastography and MRE demonstrated the best performance. In recent years, some studies have emphasized the importance of combining multiple NITs and adopting a stepwise screening approach to identify high-risk NASH, suggesting that relying on the diagnostic efficacy of any single NIT alone may be insufficient to meet clinical needs.^{32,33} We are trying to reveal the diagnostic performance of each test, providing specific data to optimize screening strategies.

Our study demonstrated a pooled AUC of 0.86 (95% CI 0.83 to 0.89) for CK18-M30 in diagnosing NASH. This is comparable to the AUC of 0.82 (95% CI 0.76 to 0.88) reported in a meta-analysis of previous studies. The CK18 fragment has been recognized as a marker for NASH in several research studies. However, due to the absence of a definitive threshold and the continual emergence of new NITs, CK18 is not currently recommended as a standard diagnostic tool in clinical practice guidelines, and its potential as a diagnostic indicator has yet to be thoroughly validated.

This study extensively examines the diagnostic accuracy of NAFLD using serological and imaging NITs, identifying a wide range of potential non-invasive diagnostic indicators. Combining multiple NITs with multi-step diagnosis is a hot topic in the diagnosis of NAFLD. Clarifying the diagnostic value of individual NITs contributes to the advancement of research in this area. We recognize that the strength of this evidence presented needs to be stronger. There are several limitations to this systematic review: (1) Many NITs rely on a single study, which may impact the statistical results and introduce confounding bias. (2) Our search was limited to studies published in Chinese and English databases, potentially leading to information bias due to excluding studies in other languages. (3) Errors in fibrosis staging often occur during clinical diagnoses, which may also contribute to information bias. (4) The included studies exhibited high heterogeneity. Through meta-regression and subgroup analysis, we found that this heterogeneity may be attributable to the design or quality of the studies, variations in the design of each original study (such as differing threshold settings), differences in the study populations, and inconsistent performance indicators across the studies.

Conclusion

The increasing number of non-invasive tests for diagnosing NAFLD creates challenges in their selective use. To help clinicians make informed decisions, we conducted a network meta-analysis ranking the accuracy for these tests. Despite the promising results we report here, not all NITs demonstrate strong accuracy, and further validation with larger datasets is necessary. Additionally, methods such as threshold calibration, cross-study validation, and clinical decision curve analysis can help clarify the benefits and risks associated with different thresholds in clinical decision-making. This approach aims to optimize clear thresholds that will guide clinical decisions and encourage research on threshold models to determine treatment thresholds, test-treatment thresholds, and testing thresholds. Future studies should focus on key aspects such as strengthening methodological reporting and providing more reliable evidence for accurate diagnosis. It is essential to ensure that the cases included in the studies are either consecutive or randomly selected. The diagnostic threshold for NITs should be established before conducting the study, rather than determining it post-analysis or leaving it unclear. Moreover, those interpreting the results should avoid being influenced by knowledge of gold standard results and should implement blinding in their analyses.

Funding

The research work was supported by Shenzhen Science and Technology Program (ZDSYS20210623092000002), Guangdong Fundamental and Applied Basic Research Program (2022B1515120034).

Disclosure

The authors report no conflicts of interest in this work.

References

- Boyle M, Tiniakos D, Schattenberg JM, et al. Performance of the PRO-C3 collagen neo-epitope biomarker in non-alcoholic fatty liver disease. *JHEP Rep*. 2019;1(3):188–198. doi:10.1016/j.jhepr.2019.06.004
- Zhou J, Zhou F, Wang W, et al. Epidemiological features of NAFLD from 1999 to 2018 in China. *Hepatology*. 2020;71(5):1851–1864. doi:10.1002/hep.31150
- Man S, Deng Y, Ma Y, et al. Prevalence of liver steatosis and fibrosis in the general population and various high-risk populations: a nationwide study with 5.7 million adults in China. *Gastroenterology*. 2023;165(4):1025–1040. doi:10.1053/j.gastro.2023.05.053
- Wang JL, Jiang SW, Hu AR, et al. Non-invasive diagnosis of non-alcoholic fatty liver disease: current status and future perspective. *Heliyon*. 2024;10(5):e27325. doi:10.1016/j.heliyon.2024.e27325
- Sumida Y, Yoneda M, Hyogo H, et al. Validation of the FIB4 index in a Japanese nonalcoholic fatty liver disease population. *BMC Gastroenterol*. 2012;12:2. doi:10.1186/1471-230X-12-2
- Cusi K, Isaacs S, Barb D, et al. American Association of Clinical Endocrinology Clinical Practice Guideline for the diagnosis and management of nonalcoholic fatty liver disease in primary care and endocrinology clinical settings: co-sponsored by the American Association for the Study of Liver Diseases (AASLD). *Endocr Pract*. 2022;28(5):528–562. doi:10.1016/j.eprac.2022.03.010
- Fouad Y, Esmat G, Elwakil R, et al. The Egyptian clinical practice guidelines for the diagnosis and management of metabolic associated fatty liver disease. *Saudi J Gastroenterol*. 2022;28(1):3–20. doi:10.4103/sjg.sjg_357_21
- Kanwal F, Shubrook JH, Adams LA, et al. Clinical care pathway for the risk stratification and management of patients with nonalcoholic fatty liver disease. *Gastroenterology*. 2021;161(5):1657–1669. doi:10.1053/j.gastro.2021.07.049
- Kang SH, Lee HW, Yoo JJ, et al. KASL clinical practice guidelines: management of nonalcoholic fatty liver disease. *Clin Mol Hepatol*. 2021;27(3):363–401. doi:10.3350/cmh.2021.0178
- Francque S, Lanthier N, Verbeke L, et al. The Belgian association for study of the liver guidance document on the management of adult and paediatric non-alcoholic fatty liver disease. *Acta Gastroenterol Belg*. 2018;81(1):55–81.
- Chalasanani N, Younossi Z, Lavine JE, et al. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American association for the study of liver diseases. *Hepatology*. 2018;67(1):328–357. doi:10.1002/hep.29367
- Gruneau L, Kechagias S, Sandström P, Ekstedt M, Henriksson M. Cost-effectiveness analysis of noninvasive tests to identify advanced fibrosis in non-alcoholic fatty liver disease. *Hepatol Commun*. 2023;7(7). doi:10.1097/HC9.0000000000000191
- Li G, Zhang X, Lin H, Liang LY, Wong GL, Wong VW. Non-invasive tests of non-alcoholic fatty liver disease. *Chin Med J*. 2022;135(5):532–546. doi:10.1097/CM9.0000000000000207
- Tilg H, Adolph TE, Moschen AR. Multiple parallel hits hypothesis in nonalcoholic fatty liver disease: revisited after a decade. *Hepatology*. 2021;73(2):833–842. doi:10.1002/hep.31518
- Thiele MS, Péloquin S, Valenti LN, et al. Assessing the applicability of non-invasive diagnostic tests (NITs) in non-alcoholic fatty liver disease: an international qualitative study. *J Hepatol*. 2022;77:S438–S439. doi:10.1016/S0168-8278(22)01213-2
- Chow KW, Futela P, Saharan A, Saab S. Comparison of guidelines for the screening, diagnosis, and noninvasive assessment of nonalcoholic fatty liver disease. *J Clin Exp Hepatol*. 2023;13(5):783–793. doi:10.1016/j.jceh.2023.01.016
- Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med*. 2015;162(11):777–784. doi:10.7326/M14-2385

18. McInnes MDF, Moher D, Thombs BD, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA*. 2018;319(4):388–396. doi:10.1001/jama.2017.19163
19. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529–536. doi:10.7326/0003-4819-155-8-201110180-00009
20. Nyaga VN, Aerts M, Arbyn M. ANOVA model for network meta-analysis of diagnostic test accuracy data. *Stat Methods Med Res*. 2018;27(6):1766–1784. doi:10.1177/0962280216669182
21. Wu JL, Pan B, Ge L, Zhang JH, Yuan CZ, Tian JH. ANOVA model for Bayesian network meta-analysis of diagnostic test accuracy. *Chinese J of Evidence-Based Med*. 2017;17(09):1111–1116.
22. Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol*. 2003;56(11):1129–1135. doi:10.1016/S0895-4356(03)00177-X
23. Desmet VJ, Gerber M, Hoofnagle JH, Manns M, Scheuer PJ. Classification of chronic hepatitis: diagnosis, grading and staging. *Hepatology*. 1994;19(6):1513–1520. doi:10.1002/hep.1840190629
24. Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR cooperative study group. *Hepatology*. 1996;24(2):289–293. doi:10.1002/hep.510240201
25. Ishak K, Baptista A, Bianchi L, et al. Histological grading and staging of chronic hepatitis. *J Hepatol*. 1995;22(6):696–699. doi:10.1016/0168-8278(95)80226-6
26. 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(5):1797–1835. doi:10.1097/HEP.0000000000000323
27. Yokoo T, Bydder M, Hamilton G, et al. Nonalcoholic fatty liver disease: diagnostic and fat-grading accuracy of low-flip-angle multiecho gradient-recalled-echo MR imaging at 1.5 T. *Radiology*. 2009;251(1):67–76. doi:10.1148/radiol.2511080666
28. Zhang X, Wong GL, Wong VW. Application of transient elastography in nonalcoholic fatty liver disease. *Clin Mol Hepatol*. 2020;26(2):128–141. doi:10.3350/cmh.2019.0001n
29. Gao Q, Ding JP, Wang FY. Advances in quantitative diagnostic imaging of nonalcoholic fatty liver disease. *Int J Med Radiol*. 2018;41(04):436–439.
30. Kouvari M, Valenzuela-Vallejo L, Guatibonza-Garcia V, et al. Liver biopsy-based validation, confirmation and comparison of the diagnostic performance of established and novel non-invasive steatotic liver disease indexes: results from a large multi-center study. *Metabolism*. 2023;147:155666. doi:10.1016/j.metabol.2023.155666
31. Shah AG, Lydecker A, Murray K, Tetri BN, Contos MJ, Sanyal AJ. Comparison of noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol*. 2009;7(10):1104–1112. doi:10.1016/j.cgh.2009.05.033
32. Younossi Z, Alkhoury N, Cusi K, et al. A practical use of noninvasive tests in clinical practice to identify high-risk patients with nonalcoholic steatohepatitis. *Aliment Pharmacol Ther*. 2023;57(3):304–312. doi:10.1111/apt.17346
33. Younossi ZM, Pham H, Felix S, et al. Identification of high-risk patients with nonalcoholic fatty liver disease using noninvasive tests from primary care and endocrinology real-world practices. *Clin Transl Gastroenterol*. 2021;12(4):e00340. doi:10.14309/ctg.0000000000000340

Clinical Epidemiology

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification, systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

Dovepress
Taylor & Francis Group