Diagn Interv Radiol 2026; DOI: 10.4274/dir.2026.264497

ABDOMINAL IMAGING

META-ANALYSIS

©Copyright 2027 Author(s) - Available online at dirjournal.org.
Content of this journal is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Magnetic resonance imaging proton density fat fraction and selected quantitative ultrasound methods for hepatic steatosis: histology-referenced diagnostic performance and longitudinal responsiveness—a systematic review and meta-analysis

Özlem Öner[1]
Mecit Kantarcı[2]

[1]Dokuz Eylül University Faculty of Medicine, Department of Anesthesiology and Reanimation, Subdivision of Critical Care Medicine, İzmir, Türkiye

[2]Atatürk University Faculty of Medicine, Department of Radiology, Erzurum, Türkiye

PURPOSE

To synthesize histology-referenced diagnostic performance at study-selected thresholds and the longitudinal responsiveness of magnetic resonance imaging proton density fat fraction (MRI-PDFF) and selected quantitative ultrasound methods in adults with non-alcoholic fatty liver disease/metabolic dysfunction-associated steatotic liver disease.

METHODS

Four databases were searched from database inception through July 1, 2026. Sensitivity, specificity, and odds ratios were pooled with 95% confidence intervals using heterogeneity-selected fixed- or random-effects models. Small-study effects were assessed using funnel plots and Begg tests; longitudinal changes were pooled when change-score standard deviations were reported.

RESULTS

In total, 44 articles were included: 34 contributed histology-referenced diagnostic data, and 11 were assessed for longitudinal responsiveness, with 1 article contributing to both. At steatosis grade (S) ≥ 1/S ≥ 2/S ≥3, MRI-PDFF sensitivity and specificity was 93.6%/81.5%/83.1% and 87.1%/86.7%/83.2%, respectively; the controlled attenuation parameter (CAP) was 86.8%/81.2%/86.4% and 98.9%/71.0%/64.6%, respectively. The attenuation coefficient/attenuation imaging sensitivity/specificity was 83.2%/98.1%, 86.4%/79.9%, and 99.3%/76.4%, respectively. Ultrasound-guided attenuation parameter and ultrasound-derived fat fraction evidence was limited. Three head-to-head longitudinal studies included 298 participants, of whom only 158 had analyzable change-score data; in these, pooled changes were −4.53 percentage points for MRI-PDFF and −12.71 dB/m for the CAP. Thirty-two of the 34 diagnostic studies were at high-risk of bias from cohort-selected thresholds. Begg tests did not detect significant small-study effects, although their power was very limited in analyses that included only 2 or 3 studies.

CONCLUSION

At study-selected thresholds, MRI-PDFF showed balanced sensitivity and specificity across steatosis grades, whereas quantitative ultrasound varied by method, platform, and threshold. These findings support comparative synthesis but do not establish validated cut-offs or superiority of either modality.

CLINICAL SIGNIFICANCE

Study-selected thresholds limit clinical translation. MRI-PDFF proton density fat fraction and quantitative ultrasound are complementary, with use depending on clinical purpose, platform, and precision needs.

KEYWORDS

Hepatic steatosis, magnetic resonance imaging, quantitative ultrasound, diagnostic accuracy, meta-analysis

Handling editor: Sonay Aydın

Corresponding author: Mecit Kantarcı

E-mail: akkanrad@hotmail.com

Received 21 September 2026; revision requested 29 September 2026; accepted 07 October 2026.

Epub: 03.11.2026

DOI: 10.4274/dir.2026.264497

You may cite this article as: Öner Ö, Kantarcı M. Magnetic resonance imaging proton density fat fraction and selected quantitative ultrasound methods for hepatic steatosis: histology-referenced diagnostic performance and longitudinal responsiveness—a systematic review and meta-analysis. *Diagn Interv Radiol*. 03 November 2026 DOI: 10.4274/dir.2026.264497 [Epub Ahead of Print].

The term non-alcoholic fatty liver disease (NAFLD) has historically been used to refer to hepatic steatosis occurring in the absence of substantial alcohol consumption and another dominant cause of liver fat accumulation. Under the updated multisociety nomenclature, metabolic dysfunction-associated steatotic liver disease (MASLD) is defined by hepatic steatosis together with at least one cardiometabolic risk factor.[1] Its clinical spectrum extends from isolated steatosis to metabolic dysfunction-associated steatohepatitis (MASH), progressive fibrosis, cirrhosis, and hepatocellular carcinoma.

Liver biopsy remains the reference standard for grading steatosis and assessing steatohepatitis and fibrosis. However, its invasive nature, sampling variability, interobserver disagreement, and limited suitability for repeated assessment have encouraged the development of non-invasive imaging biomarkers.[1] Quantitative ultrasound methods aim to improve objectivity by measuring tissue attenuation or related acoustic properties. These methods include the controlled attenuation parameter (CAP), attenuation coefficient (AC) or attenuation imaging (ATI), ultrasound-guided attenuation parameter (UGAP), and ultrasound-derived fat fraction (UDFF).[2-4]

Chemical shift–encoded magnetic resonance imaging proton density fat fraction (MRI-PDFF) corrects for major technical confounders affecting magnetic resonance signal intensity and provides a continuous estimate of hepatic fat over a substantially larger sampling volume than liver biopsy.[5] Emerging quantitative ultrasound methods provide accessible estimates of tissue attenuation or related acoustic properties, although their reliability may depend on probe selection, measurement depth, region-of-interest placement, and platform-specific quality criteria.[2-4] The diagnostic performance of both MRI-PDFF and quantitative ultrasound is also influenced by the selected operating threshold and the population from which that threshold was derived.[6,7]

Main points

- All diagnostic estimates reflect study-selected thresholds; they should not be interpreted as externally validated clinical cut-offs.
- Magnetic resonance imaging proton density fat fraction (MRI-PDFF) showed comparatively balanced sensitivity and specificity across histological grades (S), whereas the controlled attenuation parameter (CAP) specificity was lower for S ≥ 2 and S ≥ 3.
- Attenuation coefficient and attenuation imaging results varied by platform and threshold; sparse ultrasound-guided attenuation parameter and ultrasound-derived fat fraction evidence limits firm modality comparisons.
- Longitudinal CAP and MRI-PDFF changes were associated but not interchangeable, and the available paired change-score evidence was limited.

Previous meta-analyses have evaluated MRI-PDFF, CAP, and AC methods against histology. Wu et al.[8] reported areas under the curve (AUCs) at steatosis grade (S) ≥1, S ≥ 2, and S ≥ 3 of 0.99, 0.89, and 0.90, respectively, for MRI-PDFF; 0.95, 0.84, and 0.77, respectively, for the CAP; and 0.97, 0.90, and 0.89, respectively, for the AC. That review focused on diagnostic accuracy and used a hierarchical meta-analytic model. The present review adds a paired longitudinal component and reports each modality's sensitivity, specificity, and diagnostic odds ratio (DOR) separately using a different analytic approach.

This review asks whether diagnostic thresholds were set independently before testing or chosen afterward to best fit each study's own data—a distinction that matters because a threshold selected from the same data it is tested on tends to appear more favorable than it would in an independent population. We treated threshold prespecification as the central question, synthesizing the histology-referenced diagnostic performance of MRI-PDFF and selected quantitative ultrasound methods separately for each imaging method and steatosis grade and examining direct head-to-head longitudinal studies to assess how closely quantitative ultrasound changes track MRI-PDFF changes. Because thresholds were largely selected by each study and ultrasound platforms differ technically, our findings should be read as evidence specific to each method, not as universal clinical cut-offs or proof of superiority.

Methods

Study design and registration

This systematic review and meta-analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) and PRISMA for Diagnostic Test Accuracy guidelines (Supplementary Appendix 1). The protocol was prospectively registered in PROSPERO (https://www.crd.york.ac.uk/PROSPERO/view/CRD420261366403). The present review focused on MRI-PDFF and the eligible quantitative ultrasound modalities, including the CAP, Canon ATI and other AC implementations (together, AC/ATI), the UGAP, and the UDFF. Diagnostic performance was evaluated separately for each imaging modality and histological steatosis threshold. Sensitivity, specificity, and DORs were pooled using statistical models selected according to the degree of between-study heterogeneity.

In terms of terminology, because many of the included studies predated the multisociety nomenclature, the original NAFLD/non-alcoholic steatohepatitis (NASH) terminology is retained when referring to study populations as defined in the primary reports, whereas MASLD/MASH terminology is used for contemporary interpretation.[1] Throughout the manuscript, "quantitative ultrasound methods" refers collectively to the CAP, AC/ATI, UGAP, and UDFF; "AC/ATI" denotes the pooled analysis group comprising Canon ATI and other explicitly described AC implementations; and "Canon ATI" is used only when referring to that vendor-specific technology.

The PubMed/MEDLINE, Embase, Web of Science, and Cochrane Library databases were systematically searched from database inception through July 1, 2026. No language restrictions were applied. The search strategies combined terms related to the target conditions—NAFLD, MASLD, NASH, and hepatic steatosis—with modality-specific terms for MRI-PDFF, chemical shift–encoded MRI, the CAP, AC-based methods, ATI, the UGAP, and UDFF. The complete database-specific search strategies are presented verbatim in Supplementary Table 1.

Eligibility criteria

For diagnostic accuracy, studies were considered eligible if they had enrolled adults with NAFLD/MASLD/NASH; evaluated MRI-PDFF or one of the selected quantitative ultrasound methods (CAP, Canon ATI, other explicitly described AC implementations, UGAP, or UDFF); used liver histology as the reference standard; and provided or permitted reconstruction of 2 × 2 data for S ≥ 1, S ≥ 2, or S ≥ 3. Canon ATI was retained as a distinct technology in study-level extraction; it was not assumed to be interchangeable with other vendors' attenuation methods. Conventional qualitative B-mode ultrasound, non-PDFF MRI methods, and elastography without a quantitative fat measure were outside the final review scope.

The technological boundary of the quantitative ultrasound synthesis was defined prospectively around AC-based techniques (CAP, Canon ATI, and other explicitly described AC implementations), the UGAP, and the combined attenuation/backscatter-based UDFF because these were the vendor-specific quantitative ultrasound biomarkers with a sufficient histology-referenced diagnostic literature to support threshold prespecification assessment and diagnostic accuracy synthesis at the time of the search (Supplementary Table 1). Other vendor-specific AC implementations (for example, tissue ATI, ultrasound attenuation, liver fat quantification, and attenuation measurement) were eligible under the same category if they reported histology-referenced 2 × 2 data, but none meeting the eligibility criteria were identified. Backscatter coefficient, speed-of-sound-based, and other combined attenuation/backscatter fat fraction techniques (for example, USFF, UGFF) were not searched for or included; their absence should not be interpreted as a comprehensive evaluation of all quantitative ultrasound categories, and the present UDFF findings should not be generalized to these platforms. The title and Abstract accordingly refer to the specific techniques evaluated rather than to the full range of available technologies.

For the longitudinal component, eligible studies included adults with NAFLD/MASLD/NASH who underwent repeated quantitative imaging after an intervention or during follow-up. Quantitative head-to-head pooling required paired MRI-PDFF and CAP results and reporting of both mean change and its standard deviation.

Exclusion criteria

Pediatric research, animal research, reviews, editorials, conference abstracts, case reports, and protocol publications were excluded. Studies lacking histological confirmation, mixed liver disease cohorts without separable NAFLD/MASLD data, conventional qualitative ultrasound without quantitative measurements, fibrosis-only studies, insufficient diagnostic data, and duplicate or overlapping cohorts were excluded from the diagnostic meta-analysis.

Study selection and data extraction

After duplicate removal, titles and abstracts were screened against the eligibility criteria, followed by full-text assessment. A standardized form captured the study design, population, manufacturer and platform, index test threshold and its derivation, histological target, and true-positive (TP), false-positive (FP), false-negative (FN), and true-negative (TN) counts.

Outcomes

For the diagnostic accuracy meta-analysis, the primary outcomes were sensitivity and specificity for S ≥ 1, S ≥ 2, and S ≥ 3. Secondary outcomes included DORs, positive and negative likelihood ratios, and summary receiver operating characteristic (ROC) displays. For the longitudinal component, outcomes were the direction and magnitude of MRI-PDFF and CAP change, within-patient correlations between change scores, categorical agreement, and ROC performance for MRI-PDFF-defined improvement or deterioration. Standardized response means were calculated descriptively as mean change divided by the standard deviation of change when both quantities were reported.

Risk of bias and methodological quality

For the diagnostic accuracy meta-analysis, methodological quality was assessed using the revised Quality Assessment of Diagnostic Accuracy Studies tool, QUADAS-2, across the domains of patient selection, index test, reference standard, and flow and timing.[9] The two distinct Kuroda reports were assessed separately. For studies evaluating more than one eligible index modality, shared study-level domains were assessed once, whereas the index test domain was evaluated using the threshold selection information relevant to each modality. Longitudinal studies were not assessed using QUADAS-2 because the measurement property of interest was responsiveness. Instead, their methodological quality was evaluated using Box 10 of the COSMIN Risk of Bias checklist.[10]

Thresholds prespecified before analysis or adopted without modification from an independent external source were coded as "Yes." Thresholds selected or adjusted within the analyzed cohort by maximizing the Youden index, sensitivity plus specificity, target specificity, or another ROC criterion were coded as "No" and resulted in a high-risk judgment for the QUADAS-2 index test domain. When the method used to derive the threshold was insufficiently described, the signaling question was coded as "Unclear."

Statistical analysis

All tests were two-sided, and $P < 0.05$ was considered statistically significant. Analyses were performed using MedCalc Statistical Software version 23.0.2 (MedCalc Software Ltd, Ostend, Belgium). A Python environment (Python Software Foundation, Wilmington, DE, USA) was used only for data organization and figure preparation. Study-level AUC values and their corresponding *P* values were extracted as reported in the primary studies. The AUCs were summarized descriptively as sample size–weighted means and study-level ranges; they were not pooled with a hierarchical or summary ROC model. These values are therefore referred to as sample size–weighted descriptive mean AUCs throughout and should not be interpreted as pooled summary AUC estimates.

Potential small-study effects were assessed using funnel-plot inspection and Begg's rank-correlation test. Kendall's tau and the corresponding two-sided *P* value were reported. Because several modality-threshold groups contained few studies, non-significant tests were interpreted as absence of detected asymmetry rather than evidence that publication bias was absent.

For longitudinal analyses, the effect measure was the mean change from baseline. Negative values indicated a decrease during follow-up. Quantitative pooling was restricted to studies or study rows reporting both mean change and the standard deviation of change; missing change-score standard deviations were not imputed. Changes in MRI-PDFF and the CAP were pooled separately with 95% confidence intervals (CIs). Study-level Pearson correlation coefficients, Cohen's kappa coefficients, and their reported p-values were extracted for direct head-to-head comparisons. Baseline/follow-up data without an analyzable change-score variance were summarized descriptively.

Sensitivity, specificity, and DORs were pooled separately with 95% CIs. Between-study heterogeneity was evaluated with Cochran's Q and I^2. A fixed-effects model was used when Q-test $P \geq 0.10$ and $I^2 < 50\%$; otherwise, a random-effects model was used. A component-wise approach, rather than a bivariate random-effects model, was used uniformly across modality-threshold groups. With as few as two to three studies in several groups (UGAP at all three thresholds, UDFF at S ≥ 1, and the MRI-PDFF analysis subsets at S ≥ 1 and S ≥ 2), the between-study sensitivity–specificity correlation required by a bivariate model is typically unstable or non-identifiable, with estimates commonly approaching the boundary of the parameter space ($\rho \to \pm 1$); the component-wise approach avoids this instability. This approach

Table 1. Characteristics of the included diagnostic and longitudinal studies

Study	Country	Design/population	n	Modality	Reference standard	Threshold(s)/role
Kang et al.[12]	South Korea	Cross-sectional diagnostic accuracy; adults undergoing biopsy	56	MRI-PDFF	Liver histology	S≥1
Idilman et al.[13]	Türkiye	Prospective diagnostic comparison; suspected NAFLD/CLD	70	MRI-PDFF	Liver histology	S≥2
Tang et al.[14]	USA	Prospective diagnostic accuracy; biopsy-proven NAFLD	89	MRI-PDFF	Liver histology	S≥1, S≥2, S≥3
Idilman et al.[15]	Türkiye	Diagnostic comparison; hepatic steatosis evaluation	19	MRI-PDFF	Liver histology	S≥2
Middleton et al.[16]	USA	Prospective multicenter study; NASH CRN population	113	MRI-PDFF	Central liver histology	S≥2, S≥3
Paige et al.[17]	USA	Prospectively designed cross-sectional pilot study	60	MRI-PDFF, investigational QUS-AC	Liver histology	MRI-PDFF: S≥2, S≥3; investigational QUS-AC: S≥2, S≥3
Park et al.[18]	South Korea	Prospective diagnostic comparison; suspected NAFLD	78	MRI-PDFF, CAP	Liver histology	MRI-PDFF: S≥1, S≥2, S≥3; CAP: S≥1, S≥2, S≥3
Beyer et al.[19]	Germany	Diagnostic cohort with paired imaging and biopsy	451*	MRI-PDFF, CAP	Liver histology	MRI-PDFF: S≥1, S≥2, S≥3; CAP: S≥1, S≥2, S≥3
Shao et al.[11]	China	Prospective cohort with cross-sectional and longitudinal components	102 209 158*	MRI-PDFF, CAP	Histology for diagnosis; MRI-PDFF for monitoring; histology/MRI-PDFF	MRI-PDFF: S≥1, S≥2, S≥3; CAP: S≥1, S≥2, S≥3
Kim et al.[20]	South Korea	Diagnostic cohort; biopsy-proven NAFLD	47	MRI-PDFF	Liver histology	S≥2, S≥3
Nogami et al.[21]	Japan	Retrospective cross-sectional study; biopsy-proven NAFLD	163	MRI-PDFF, CAP	Liver histology	MRI-PDFF: S≥1, S≥2, S≥3; CAP: S≥1, S≥2, S≥3
Gottfriedova et al.[22]	Czech Republic	Diagnostic cohort; hepatic fat assessment	56	MRI-PDFF	Liver histology	S≥1, S≥2
Runge et al.[23]	Netherlands	Prospective diagnostic comparison	55	CAP	Liver histology	S≥1, S≥2, S≥3
Chan et al.[6]	Hong Kong, China	Prospective diagnostic study; biopsy-assessed NAFLD	101	CAP	Liver histology	S≥2, S≥3
de Lédinghen et al.[24]	France	Prospective multicenter diagnostic study	261	CAP	Liver histology	S≥2, S≥3
Lee et al.[25]	South Korea	Retrospective cohort; suspected NAFLD	183	CAP	Liver histology	S≥1, S≥2, S≥3
Naveau et al.[26]	France	Prospective bariatric surgery cohort	123	CAP	Intraoperative biopsy	S≥1, S≥2, S≥3
Eddowes et al.[7]	United Kingdom	Prospective multicenter NAFLD cohort	380	CAP	Liver histology	S≥1, S≥2, S≥3
Shalimar et al.[27]	India	Prospective diagnostic study; NAFLD	219	CAP	Liver histology	S≥1, S≥2, S≥3
Huang et al.[28]	China	Prospective single-center diagnostic study	60	CAP, AC/ATI	Liver histology	CAP: S≥1, S≥2, S≥3; AC/ATI: S≥1, S≥2, S≥3
Zhou et al.[29]	China	Diagnostic comparison; biopsy-proven NAFLD	73	CAP, Canon ATI	Liver histology	CAP: S≥1, S≥2, S≥3; Canon ATI: S≥1, S≥2, S≥3
Seo et al.[30]	South Korea	Diagnostic comparison; adults undergoing biopsy	105	CAP, AC/ATI	Liver histology	CAP: S≥1, S≥2, S≥3; AC/ATI: S≥1, S≥2, S≥3
Hu et al.[31]	China	Prospective sleeve-gastrectomy cohort; baseline biopsy-referenced diagnostic component and 6-month longitudinal follow-up	160	CAP	Intraoperative biopsy	Diagnostic: S≥1, S≥2, S≥3; longitudinal: preoperative to 6 months
Tada et al.[2]	Japan	Diagnostic cohort; chronic liver disease	38	AC/ATI	Liver histology	S≥1, S≥2, S≥3

Table 1. Continued

Study	Country	Design/population	n	Modality	Reference standard	Threshold(s)/role
Sugimoto et al.[32]	Japan	Prospective diagnostic study	111	AC/ATI	Liver histology	S≥1, S≥2, S≥3
Dioguardi Burgio et al.[33]	France	Diagnostic cohort; adults undergoing biopsy	187	Canon ATI	Liver histology	S≥1, S≥2
Kuroda et al.[3]	Japan	Prospective diagnostic comparison	125	Canon ATI, UGAP	Liver histology	Canon ATI: S≥1, S≥2, S≥3; UGAP: S≥1, S≥2, S≥3
Kuroda et al.[34]	Japan	Prospective diagnostic comparison	202	UGAP	Same-day liver biopsy	S≥2, S≥3
Lee et al.[35]	South Korea	Prospective biopsy-referenced diagnostic study	102	AC/ATI	Liver histology	S≥1, S≥2, S≥3
Ogino et al.[36]	Japan	Diagnostic comparison	84	UGAP	Liver histology	UGAP: S≥1, S≥2, S≥3
Jang et al.[37]	South Korea	Diagnostic study; adults undergoing biopsy	132	Canon ATI	Liver histology	S≥1, S≥2, S≥3
Cannella et al.[38]	Italy/France	Diagnostic cohort	100	UGAP	Liver histology	S≥2, S≥3
Labyed and Milkowski[4]	USA	Prospective biopsy-referenced diagnostic study	90	UDFF	Liver histology	S≥1, S≥2, S≥3
Nakamura et al.[39]	Japan	Diagnostic study	73	UDFF	Liver histology	S≥1, S≥2, S≥3
Study	**Country**	**Design/population**	**Baseline/ follow-up n**	**Modality**	**Reference/ comparator**	**Follow-up/outcome**
An et al.[40]	China	Prospective/non-randomized treatment-cohort analysis	185 total; 95 longitudinal dataset	CAP and MRI-PDFF	MRI-PDFF	24 weeks
Wang et al.[41]	Taiwan	Prospective consecutive cohort	50/45	CAP and MRI-PDFF	MRI-PDFF	Approximately 12 months
Agarwal et al.[42]	India	Prospective bariatric surgery cohort	58/58	CAP and LSM	Paired liver biopsy	Baseline–1 year
Nixdorf et al.[43]	Austria	Prospective bariatric surgery cohort	93 at baseline; follow-up count reported in the primary study	CAP, VCTE, and 2D-SWE	Baseline intraoperative biopsy	Baseline–3 months
Hu et al.[44]	China	Prospective bariatric surgery cohort	160 at baseline; 120 at follow-up	UDFF	Intraoperative biopsy	Preoperative–postoperative follow-up
Noureddin et al.[45]	USA	Secondary analysis of randomized trial; biopsy-proven NAFLD	50	MRI-PDFF	Paired liver biopsy	Baseline–24 weeks
Loomba et al.[46]	Multicenter	Randomized phase II trial; NASH F2–F3	72	MRI-PDFF	Paired liver biopsy	24 weeks
Loomba et al.[47]	USA, multicenter	Secondary analysis of FLINT trial	78	MRI-PDFF	Paired histology	≥30% MRI-PDFF decline vs. histological response
Huang et al.[48]	USA	Prospective longitudinal NAFLD cohort	108	MRI-PDFF and ALT	Paired liver biopsy	Histological response/ NASH resolution
Argo et al.[49]	USA	Double-blind randomized placebo-controlled NASH trial	34 completed	MRI-based liver fat assessment	Paired liver biopsy	Narrative/excluded from change-score pooling

*For Beyer et al.,[19] the pooled cohort comprised 581 participants overall; 451 had baseline liver biopsy results and constitute the biopsy-referenced diagnostic denominator used in this histology-referenced meta-analysis. For Shao et al.,[11] 102 participants contributed to the diagnostic analysis, 209 were enrolled overall, and 158 were included in the longitudinal analyses. AC, attenuation coefficient; ATI, attenuation imaging; CAP, controlled attenuation parameter; CLD, chronic liver disease; MRI-PDFF, magnetic resonance imaging proton density fat fraction; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; QUS-AC, quantitative ultrasound attenuation coefficient; UGAP, ultrasound-guided attenuation parameter; UDFF, ultrasound-derived fat fraction.

does not preserve the within-study sensitivity–specificity correlation and therefore does not provide a summary ROC operating point; results were interpreted accordingly. A sensitivity analysis restricted to studies using prespecified or independently validated thresholds (QUADAS-2 index test domain rated low risk) was planned provided that a sufficient number of eligible studies were available for a given modality-threshold group.

Diagnostic accuracy analyses were performed separately by imaging modality and histological threshold: S0 vs. S1–S3 (S ≥ 1), S0–S1 vs. S2–S3 (S ≥ 2), and S0–S2 vs. S3 (S ≥ 3). TP, FP, FN, and TN counts were extracted or reconstructed for each eligible study. Sensitivity was calculated as TP/(TP+FN), specificity as TN/(TN+FP), and the DOR as (TP×TN)/(FP×FN).

Results

Study selection and characteristics

The search identified 20,127 records: 4,150 from PubMed, 989 from the Cochrane Library, 10,507 from Embase, and 4,481 from Web of Science. After the removal of 7,079 duplicates, 13,048 titles and abstracts were screened and 44 articles were included. Of these, 34 contributed histology-referenced diagnostic accuracy data and 11 were assessed for longitudinal responsiveness; 1 article (Shao et al.[11]) contributed to both components, so the two groups are not mutually exclusive (Table 1). The final synthesis was limited to MRI-PDFF and the selected quantitative ultrasound methods defined in the Methods Section. Prior reviews were used for cross-checking, but study eligibility and quantitative extraction were based on the primary reports. The PRISMA flow diagram and study-level provenance notes document the selection and verification process (Figure 1). ROC plots for all diagnostic meta-analyses are consolidated in Supplementary Figure 1.

Histology-referenced magnetic resonance imaging proton density fat fraction performance

For S ≥ 1, seven studies including 995 participants were pooled. Sensitivity was 93.6% (95% CI, 87.3–96.9; I^2 = 81.2%; Q-test P < 0.001; random effects), specificity was 87.1% (95% CI, 78.1–92.8; I^2 = 0.0%; Q-test P = 0.76; fixed effects), and the DOR was 144.51 (95% CI, 55.01–379.66; I^2 = 0.0%; Q-test P = 0.65; fixed effects). Study-level AUCs ranged from 0.95 to 1.00, with a sample size–weighted descriptive mean of 0.980 (Table 2). Begg's test gave Kendall's tau = 0.43 (P = 0.24), with no statistically significant asymmetry (Table 2; Supplementary Figure 2).

For S ≥ 2, 11 studies including 1,248 participants were pooled. Sensitivity was 81.5% (95% CI, 75.8–86.2; I^2 = 52.0%; Q-test P = 0.02; random effects), specificity was 86.7% (95% CI, 80.1–91.3; I^2 = 68.1%; Q-test P < 0.001; random effects), and the DOR was 37.42 (95% CI, 17.14–81.67; I^2 = 76.2%; Q-test P < 0.001; random effects). Study-level AUCs ranged from 0.77 to 0.99, with a sample size–weighted descriptive mean of 0.874 (Table 2). Begg's Kendall's tau was 0.40 (P = 0.09), indicating no statistically significant asymmetry (Table 2; Supplementary Figure 3).

For S ≥ 3, eight studies including 1,103 participants were pooled. Sensitivity was 83.1% (95% CI, 71.5–90.6; I^2 = 50.3%; Q-test P = 0.04; random effects), specificity was 83.2% (95% CI, 78.1–87.3; I^2 = 63.5%; Q-test P = 0.008; random effects), and the DOR was 30.27 (95% CI, 13.50–67.88; I^2 = 54.7%; Q-test P = 0.03; random effects). Study-level AUCs ranged from 0.81 to 0.97, with a sample size–weighted descriptive mean of 0.877 (Table 2). Begg's Kendall's tau was 0.21 (P = 0.55), with no statistically significant asymmetry (Table 2; Supplementary Figure 4).

A restricted sensitivity analysis subset for MRI-PDFF at S ≥ 1 and S ≥ 2, limited to studies with prespecified or independently validated thresholds, is additionally reported in Tables 2 and 3, with corresponding publication bias and forest plots shown in Supplementary Figures 5 and 6, respectively.

Histology-referenced quantitative ultrasound performance

For the CAP at S ≥ 1, 14 studies including 2,253 participants were pooled. Study-level sensitivity estimates ranged from 71.8% to 98.3%; pooled sensitivity was 86.8% (95% CI, 82.6–90.0; I^2 = 81.5%; Q-test P < 0.001; random effects). Study-level specificity ranged from 60.0% to 100.0%; pooled specificity was 98.9% (95% CI, 95.1–99.8; I^2 = 91.7%; Q-test P < 0.001; random effects). Study-level DORs ranged from 13.50 to 1,071.00; the pooled DOR was 31.64 (95% CI, 20.22–49.51; I^2 = 18.6%; Q-test P = 0.251; fixed effects), and AUCs ranged from 0.770 to 0.990, with a sample size–weighted descriptive mean of 0.913. Begg's tau was 0.38 (P = 0.11) (Supplementary Figure 7).

For the CAP at S ≥ 2, 15 studies including 2,514 participants were pooled. Study-level sensitivity ranged from 63.3% to 93.8%; pooled sensitivity was 81.2% (95% CI, 76.8–85.0; I^2 = 68.5%; Q-test P < 0.001; random effects). Study-level specificity ranged from 40.9% to 86.7%; pooled specificity was 71.0% (95% CI, 63.1–77.8; I^2 = 85.4%; Q-test P < 0.001; random effects). Study-level DORs ranged from 2.47 to 91.00; the pooled DOR was 11.50 (95% CI, 7.21–18.35; I^2 = 79.7%;

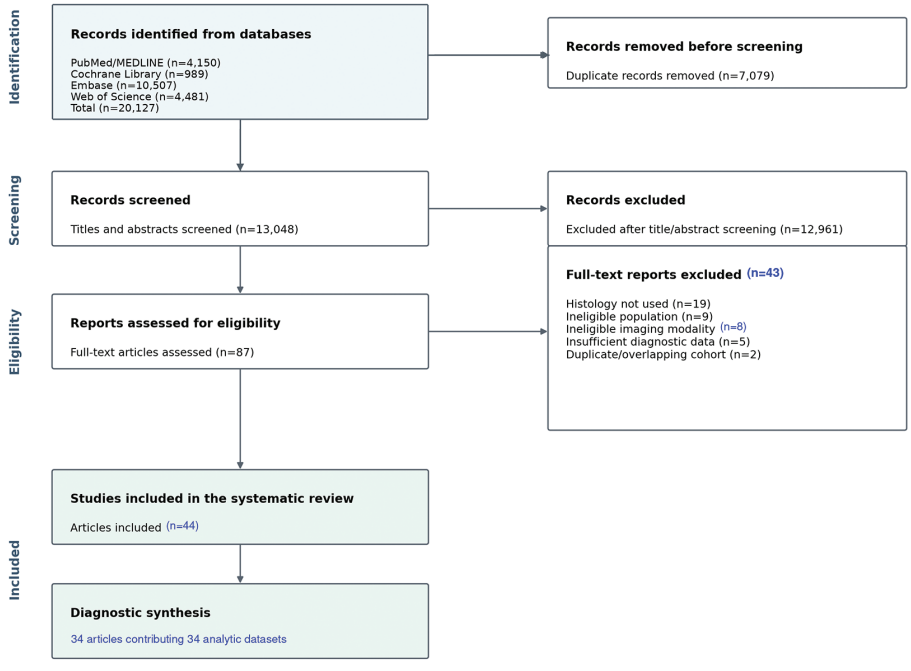


Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 flow diagram of study selection. Of the 20,127 identified records, 7,079 duplicates were removed, 13,048 records were screened, 87 full texts were assessed, 43 were excluded, and 44 articles were included.

Q-test $P < 0.001$; random effects), and AUCs ranged from 0.600 to 0.940, with a sample size–weighted descriptive mean of 0.767. Begg's tau was 0.40 ($P = 0.09$) (Supplementary Figure 8).

For the CAP at S ≥ 3, 15 studies including 2,514 participants were pooled. Study-level sensitivity ranged from 61.3% to 100.0%; pooled sensitivity was 86.4% (95% CI, 77.5–92.1; $I^2 = 85.8\%$; Q-test $P < 0.001$; random effects). Study-level specificity ranged from 47.2% to 76.1%; pooled specificity was 64.6% (95% CI, 60.2–68.7; $I^2 = 70.6\%$; Q-test $P < 0.001$; random effects). Study-level DORs ranged from 2.28 to 56.10; the pooled DOR was 7.42 (95% CI, 4.67–11.79; $I^2 = 67.6\%$; Q-test $P < 0.001$; random effects), and AUCs ranged from 0.630 to 0.870, with a sample size–weighted descriptive mean of 0.727. Begg's tau was 0.42 ($P = 0.08$) (Supplementary Figure 9).

For the AC/ATI at S ≥ 1, 11 studies including 1,219 participants were pooled. Study-level sensitivity ranged from 61.5% to 95.6%; pooled sensitivity was 83.2% (95% CI, 77.9–87.4; $I^2 = 74.0\%$; Q-test $P < 0.001$; random effects). Study-level specificity ranged from 75.0% to 100.0%; pooled specificity was 98.1% (95% CI, 93.1–99.5; $I^2 = 87.0\%$; Q-test $P < 0.001$; random effects). Study-level DORs ranged from 3.78 to 272.45; the pooled DOR was 43.03 (95% CI, 22.88–80.93; $I^2 = 17.6\%$; Q-test $P = 0.276$; fixed effects), and AUCs ranged from 0.770 to 0.980, with a sample size–weighted descriptive mean of 0.913. Begg's tau was 0.02 ($P = 0.99$) (Supplementary Figure 10).

For the AC/ATI at S ≥ 2, 12 studies including 1,279 participants were pooled. Study-level sensitivity ranged from 68.9% to 95.3%; pooled sensitivity was 86.4% (95% CI, 80.3–90.8; $I^2 = 66.1\%$; Q-test $P < 0.001$; random effects). Study-level specificity ranged from 65.6% to 91.8%; pooled specificity was 79.9% (95% CI, 74.5–84.4; $I^2 = 58.0\%$; Q-test $P = 0.006$; random effects). Study-level DORs ranged from 9.17 to 80.86; the pooled DOR was 24.47 (95% CI, 17.43–34.37; $I^2 = 37.3\%$; Q-test $P = 0.093$; fixed effects), and AUCs ranged from 0.790 to 0.950, with a sample size–weighted descriptive mean of 0.886. Begg's tau was 0.36 ($P = 0.12$) (Supplementary Figure 11).

For the AC/ATI at S ≥ 3, 11 studies including 1,092 participants were pooled. Study-level sensitivity ranged from 61.1% to 100.0%; pooled sensitivity was 99.3% (95% CI, 95.3–99.9; $I^2 = 92.1\%$; Q-test $P < 0.001$; random effects). Study-level specificity ranged from 60.5% to 84.9%; pooled specificity was 76.4% (95% CI, 70.7–81.3; $I^2 = 70.6\%$; Q-test $P < 0.001$; random effects). Study-level DORs ranged from 8.87 to 131.05; the pooled DOR was 22.91 (95% CI, 13.37–39.25; $I^2 = 0.0\%$; Q-test $P = 0.734$; fixed effects), and AUCs ranged from 0.770 to 0.940, with a sample size–weighted descriptive mean of 0.878. Begg's tau was 0.35 ($P = 0.17$) (Supplementary Figure 12).

For the UGAP at S ≥ 1, two studies including 209 participants were pooled. Study-level sensitivity ranged from 84.8% to 86.7%; pooled sensitivity was 85.5% (95% CI, 79.6–90.0; $I^2 = 0.0\%$; Q-test $P = 0.720$; fixed effects). Study-level specificity ranged from 80.0% to 88.9%; pooled specificity was 82.3% (95% CI, 63.8–92.5; $I^2 = 0.0\%$; Q-test $P = 0.563$; fixed effects). Study-level DORs ranged from 22.25 to 52.00; the pooled DOR was 27.22 (95% CI, 9.40–78.85; $I^2 = 0.0\%$; Q-test $P = 0.506$; fixed

Table 2. Pooled diagnostic performance and Begg publication bias assessment by imaging method and histologic steatosis threshold

Method	Threshold	Studies	n	Sensitivity, % (95% CI)	Specificity, % (95% CI)	DOR (95% CI)	Sample size–weighted descriptive mean AUC (range)	*P* values	Begg tau	Begg *P*
MRI-PDFF	S ≥ 1	7	995	93.6 (87.3–96.9)	87.1 (78.1–92.8)	144.51 (55.01–379.66)	0.980 (0.950–1.000)	< 0.001	0.43	0.24
MRI-PDFF	S ≥ 2	11	1.248	81.5 (75.8–86.2)	86.7 (80.1–91.3)	37.42 (17.14–81.67)	0.874 (0.770–0.990)	< 0.001	0.40	0.09
MRI-PDFF	S ≥ 3	8	1.103	83.1 (71.5–90.6)	83.2 (78.1–87.3)	30.27 (13.50–67.88)	0.877 (0.810–0.970)	< 0.001	0.21	0.55
MRI-PDFF analysis[a]	S ≥ 1	2	73	93.3 (83.5–97.5)	93.9 (66.9–99.1)	206.36 (17.78–2395.05)	0.991 (0.988–1.000)	< 0.001	0.01	0.99
MRI-PDFF analysis[a]	S ≥ 2	2	74	92.1 (78.2–97.4)	88.6 (42.3–98.8)	104.06 (17.29–626.38)	0.945 (0.857–0.975)	< 0.001	0.01	0.99
CAP	S ≥ 1	14	2.253	86.8 (82.6–90.0)	98.9 (95.1–99.8)	31.64 (20.22–49.51)	0.913 (0.770–0.990)	< 0.001	0.38	0.11
CAP	S ≥ 2	15	2.514	81.2 (76.8–85.0)	71.0 (63.1–77.8)	11.50 (7.21–18.35)	0.767 (0.600–0.940)	< 0.001	0.40	0.09
CAP	S ≥ 3	15	2.514	86.4 (77.5–92.1)	64.6 (60.2–68.7)	7.42 (4.67–11.79)	0.727 (0.630–0.870)	< 0.001	0.42	0.08
AC/ATI	S ≥ 1	11	1.219	83.2 (77.9–87.4)	98.1 (93.1–99.5)	43.03 (22.88–80.93)	0.913 (0.770–0.980)	< 0.001	0.02	0.99
AC/ATI	S ≥ 2	12	1.279	86.4 (80.3–90.8)	79.9 (74.5–84.4)	24.47 (17.43–34.37)	0.886 (0.790-0.950)	< 0.001	0.36	0.12
AC/ATI	S ≥ 3	11	1.092	99.3 (95.3–99.9)	76.4 (70.7–81.3)	22.91 (13.37–39.25)	0.878 (0.770–0.940)	< 0.001	0.35	0.17
UGAP[a]	S ≥ 1	2	209	85.5 (79.6–90.0)	82.3 (63.8–92.5)	27.22 (9.40–78.85)	0.910 (0.890–0.940)	< 0.001	0.01	0.99
UGAP[a]	S ≥ 2	3	309	79.6 (72.5–85.2)	86.9 (80.4–91.5)	28.59 (14.90–54.84)	0.893 (0.828–0.950)	< 0.001	0.33	0.99
UGAP[a]	S ≥ 3	3	309	90.0 (81.3–94.9)	76.3 (70.2–81.4)	26.28 (11.78–58.64)	0.877 (0.830–0.912)	< 0.001	0.33	0.99
UDFF[a]	S ≥ 1	2	163	89.4 (72.8–96.3)	99.5 (9.4–100.0)	81.67 (19.08–349.57)	0.947 (0.940–0.956)	< 0.001	0.01	0.99

AUC, area under the curve; CI, confidence interval; DOR, diagnostic odds ratio; Begg's tau, Kendall's tau from Begg's rank-correlation test for small-study effects; Begg's *P*, corresponding two-sided *P* value; MRI-PDFF, magnetic resonance imaging proton density fat fraction; CAP, controlled attenuation parameter; AC, attenuation coefficient; ATI, attenuation imaging; UGAP, ultrasound-guided attenuation parameter; UDFF, ultrasound-derived fat fraction. Sample size–weighted descriptive mean AUC values are descriptive summaries rather than pooled summary AUC estimates from a hierarchical or summary receiver operating characteristic model, and the values in parentheses represent the range of study-level AUCs rather than 95% CIs.

a, estimates based on only two or three studies (MRI-PDFF analysis subsets at S ≥ 1 and S ≥ 2; UGAP at S ≥ 1, S ≥ 2, and S ≥ 3; and UDFF at S ≥ 1) are highly imprecise and should be interpreted with particular caution. In particular, the pooled UDFF specificity at S ≥ 1 (95% CI, 9.4–100.0) does not indicate near-perfect specificity, and Begg's test has very limited power to detect small-study effects in these groups. All estimates reflect study-selected thresholds and should not be interpreted as externally validated cut-offs.

effects), and AUCs ranged from 0.890 to 0.940, with a sample size–weighted descriptive mean of 0.910. Begg's tau was 0.01 (P = 0.99) (Supplementary Figure 13).

For the UGAP at S ≥ 2, three studies including 309 participants were pooled. Study-level sensitivity ranged from 75.7% to 85.7%; pooled sensitivity was 79.6% (95% CI, 72.5–85.2; I^2 = 0.0%; Q-test P = 0.459; fixed effects). Study-level specificity ranged from 84.5% to 91.8%; pooled specificity was 86.9% (95% CI, 80.4–91.5; I^2 = 0.0%; Q-test P = 0.501; fixed effects). Study-level DORs ranged from 20.26 to 67.50; the pooled DOR was 28.59 (95% CI, 14.90–54.84; I^2 = 0.0%; Q-test P = 0.384; fixed effects), and AUCs ranged from 0.828 to 0.950, with a sample size–weighted descriptive mean of 0.893. Begg's tau was 0.33 (P = 0.99) (Supplementary Figure 14).

For the UGAP at S ≥ 3, three studies including 309 participants were pooled. Study-level sensitivity ranged from 85.7% to 92.3%; pooled sensitivity was 90.0% (95% CI, 81.3–94.9; I^2 = 0.0%; Q-test P = 0.800; fixed effects). Study-level specificity ranged from 67.2% to 80.0%; pooled specificity was 76.3% (95% CI, 70.2–81.4; I^2 = 47.1%; Q-test P = 0.151; fixed effects). Study-level DORs ranged from 19.50 to 47.40; the pooled DOR was 26.28 (95% CI, 11.78–58.64; I^2 = 0.0%; Q-test P = 0.657; fixed effects), and AUCs ranged from 0.830 to 0.912, with a sample size–weighted descriptive mean of 0.877. Begg's tau was 0.33 (P = 0.99) (Supplementary Figure 15).

For the UDFF at S ≥ 1, two studies including 163 participants were pooled. Study-level sensitivity ranged from 84.5% to 94.9%; pooled sensitivity was 89.4% (95% CI, 72.8–96.3; I^2 = 58.7%; Q-test P = 0.120; random effects). Study-level specificity ranged from 82.4% to 100.0%; pooled specificity was 99.5% (95% CI, 9.4–100.0; I^2 = 96.0%; Q-test P < 0.001; random effects). Study-level DORs ranged from 68.85 to 86.33; the pooled DOR was 81.67 (95% CI, 19.08–349.57; I^2 = 0.0%; Q-test P = 0.896; fixed effects), and AUCs ranged from 0.940 to 0.956, with a sample size–weighted descriptive mean of 0.947. Begg's tau was 0.01 (P = 0.99) (Supplementary Figure 16). The very wide specificity CI indicates extreme imprecision and does not support interpreting the point estimate as stable near-perfect specificity.

Only Labyed and Milkowski[4] reported UDFF data for S ≥ 2 and S ≥ 3. Because a single study was available at each threshold, no meta-analysis was performed, and these findings were retained as study-level diagnostic results only. Across quantitative ultrasound analyses, Begg test results were not statistically significant; however, results based on only two or three studies have very limited power to detect asymmetry (Table 2). Across modality-threshold groups, the available study-level tests of AUC against 0.50 were statistically significant (all reported P < 0.001; Table 2).

Longitudinal analyses

In the head-to-head longitudinal analysis, three studies including 298 participants were identified, but only three rows comprising 158 participants reported the change-score variance required for pooling and were available for each modality. The pooled MRI-PDFF

Table 3. Heterogeneity and model selection for diagnostic meta-analyses

Method	Threshold	Sens. I^2, %	P	Model	Spec. I^2, %	P	Model	DOR I^2, %	P	Model
MRI-PDFF	S ≥ 1	81.2	< 0.001	Random effects	0.0	0.76	Fixed effects	0.0	0.65	Fixed effects
MRI-PDFF	S ≥ 2	52.0	0.02	Random effects	68.1	< 0.001	Random effects	76.2	< 0.001	Random effects
MRI-PDFF	S ≥ 3	50.3	0.04	Random effects	63.5	0.008	Random effects	54.7	0.03	Random effects
MRI-PDFF analysis	S ≥ 1	0.0	0.89	Fixed effects	0.0	0.76	Fixed effects	0.0	0.74	Fixed effects
MRI-PDFF analysis	S ≥ 2	0.0	0.946	Fixed effects	70.3	0.067	Random effects	46.2	0.173	Fixed effects
CAP	S ≥ 1	81.5	< 0.001	Random effects	91.7	< 0.001	Random effects	18.6	0.251	Fixed effects
CAP	S ≥ 2	68.5	< 0.001	Random effects	85.4	< 0.001	Random effects	79.7	< 0.001	Random effects
CAP	S ≥ 3	85.8	< 0.001	Random effects	70.6	< 0.001	Random effects	67.6	< 0.001	Random effects
AC/ATI	S ≥ 1	74.0	< 0.001	Random effects	87.0	< 0.001	Random effects	17.6	0.276	Fixed effects
AC/ATI	S ≥ 2	66.1	< 0.001	Random effects	58.0	0.006	Random effects	37.3	0.093	Fixed effects
AC/ATI	S ≥ 3	92.1	< 0.001	Random effects	70.6	< 0.001	Random effects	0.0	0.734	Fixed effects
UGAP	S ≥ 1	0.0	0.720	Fixed effects	0.0	0.563	Fixed effects	0.0	0.506	Fixed effects
UGAP	S ≥ 2	0.0	0.459	Fixed effects	0.0	0.501	Fixed effects	0.0	0.384	Fixed effects
UGAP	S ≥ 3	0.0	0.80	Fixed effects	47.1	0.15	Fixed effects	0.0	0.66	Fixed effects
UDFF	S ≥ 1	58.7	0.12	Random effects	96.0	< 0.001	Random effects	0.0	0.89	Fixed effects

No row met the prespecified threshold for statistical significance (P < 0.05).
MRI-PDFF, magnetic resonance imaging proton density fat fraction; CAP, controlled attenuation parameter; AC, attenuation coefficient; ATI, attenuation imaging; UGAP, ultrasound-guided attenuation parameter; UDFF, ultrasound-derived fat fraction; DOR, diagnostic odds ratio; Sens., sensitivity; Spec., specificity. Model selection reflects the prespecified fixed-/random-effects selection rule described under "Statistical analysis".

change was −4.53 percentage points (95% CI, −6.26 to −2.80; I^2 = 67.5%; Q-test P = 0.046; random effects), and the pooled CAP change was −12.71 dB/m (95% CI, −19.66 to −5.76; I^2 = 65.5%; Q-test P = 0.055; random effects) (Table 4). Begg's Kendall's tau was 0.33 (P = 0.99) for both analyses. The head-to-head synthesis was restricted to rows reporting mean change and the standard deviation of change; missing change-score variances were not imputed. In An et al.,[40] absolute CAP and MRI-PDFF changes were correlated (r = 0.492, P < 0.0001), as were relative changes (r = 0.598, P < 0.0001); categorical change agreement was moderate (κ = 0.424, P = 0.00011), and 73.8% changed in the same direction. In Shao et al.,[11] change correlations were r = 0.612 and r = 0.524 in moderate-to-severe and mild steatosis, respectively (both P < 0.001), whereas categorical agreement was weak (κ = 0.181, P = 0.001), and AUCs were 0.702 (P = 0.003) for deterioration and 0.717 (P < 0.001) for improvement in MRI-PDFF grade. In Wang et al.,[41] CAP change correlated with MRI-PDFF change (n = 45, r = 0.493, P = 0.001); in multivariable linear regression, CAP change remained independently associated with MRI-PDFF change (β = 0.039, standard error = 0.011, P = 0.0009). Hu et al.[31] contained follow-up CAP data but lacked the paired repeated MRI-PDFF change-score information required for the head-to-head meta-analysis and was retained narratively. In the separate MRI follow-up analysis based on two Patel et al.[50] groups (total n = 35), mean change was −4.10 among histological responders and −0.60 among non-responders; the pooled mean change was −2.10 (95% CI, −5.50 to 1.29; I^2 = 74.9%; Q-test P = 0.046; random effects; Begg's tau = 0.33, P = 0.99).

Methodological quality

The QUADAS-2 assessment classified the index test domain as high-risk in 32 of the 34 diagnostic studies (94.1%) and as unclear in 2 (5.9%), largely because operating thresholds were selected within the analyzed cohorts (Figure 2). The two unclear judgments concerned Tada et al.[2] and Labyed and Milkowski[4]. Among the 11 longitudinal studies assessed for responsiveness, 6 were rated adequate and 5 doubtful; none were rated very good or inadequate overall. Detailed judgments, including the complete QUADAS-2 index test domain assessment and the COSMIN responsiveness ratings, are provided in Supplementary Tables 2–5. Because no diagnostic study met the low-risk threshold prespecification definition in any modality-threshold group, the planned sensitivity analysis restricted to studies with prespecified or independently validated thresholds could not be performed; this inability is itself an important finding since it indicates that the histology-referenced literature for these techniques does not currently contain a subset of studies from which externally validated operating characteristics could be derived. Characteristics of the included diagnostic and longitudinal studies are summarized in Table 1, and funnel and forest plots for the diagnostic and longitudinal meta-analyses are consolidated in Figures 3–6.

Discussion

Principal findings and interpretation

This review synthesized histology-referenced performance at study-selected thresholds and paired longitudinal responsiveness. The results demonstrated that MRI-PDFF estimates were comparatively balanced across grades, whereas quantitative ultrasound estimates varied by method and histological target. The evidence does not support a universal cut-off or a formal claim that one modality is superior: threshold optimism, substantial heterogeneity, sparse evidence for several ultrasound methods, and indirect cross-study contrasts constrain clinical interpretation.

Relationship to Wu et al. [8] and other prior diagnostic meta-analyses

The most directly comparable prior synthesis is that of Wu et al.,[8] who performed a bivariate random-effects meta-analysis of MRI-PDFF, CAP, and AC techniques against liver histology across the same three steatosis thresholds using PubMed, the Cochrane Library, Embase, and Web of Science. Our pooled estimates are numerically close to theirs: CAP specificity was 71.0% and 64.6% in our S ≥ 2 and S ≥ 3 analyses, compared with approximately 70% and 65% in Wu et al.;[8] CAP sensitivity at S ≥ 2 was 81.2% in our analysis, compared with approximately 82% in theirs. Our sample size–weighted descriptive mean AUCs for MRI-PDFF, CAP, and AC/ATI at S ≥ 1/S ≥ 2/S ≥ 3 (0.980/0.874/0.877; 0.913/0.767/0.727; 0.913/0.886/0.878, respectively) are likewise broadly consistent with the AUCs reported by Wu et al.[8] (0.99/0.89/0.90 for MRI-PDFF; 0.95/0.84/0.77 for CAP; 0.97/0.90/0.89 for AC). Thus, the numerical patterns were broadly consistent across the two reviews, including comparatively stable MRI-PDFF estimates across grades and lower CAP specificity at higher grades. However, because the two reviews drew on partially overlapping and therefore non-independent sets of studies and the present review did not perform a formal direct comparative meta-analysis, this consistency should not be interpreted as establishing a ranking of the techniques or as a new diagnostic accuracy finding in its own right.

The present review differs from that of Wu et al.[8] in three respects that define its distinct contribution rather than a restatement of existing knowledge. First, we performed a QUADAS-2 reassessment focused specifically on threshold prespecification, finding that 32 of the 34 diagnostic studies, across all technologies, selected their operating threshold within the same cohort used to evaluate diagnostic performance; this reassessment was not part of the methodological quality evaluation reported by Wu et al.[8] and directly qualifies how the pooled operating characteristics—ours and theirs—should be interpreted. Second, we evaluated the UGAP and UDFF, technologies not included in the review by Wu et al.,[8] and treated Canon ATI as a distinct technology from other AC implementations rather than assuming inter-

Table 4. Longitudinal change-score meta-analyses

Analysis	Rows/ groups	n	Pooled mean change (95% CI)	I^2, %	P	Model	Kendall's tau	Begg's P
MRI-PDFF head-to-head change	3	158	−4.53 (−6.26 to −2.80)	67.5	0.046	Random effects	0.33	0.99
CAP head-to-head change	3	158	−12.71 (−19.66 to −5.76)	65.5	0.055	Random effects	0.33	0.99
MRI histologic response groups	2	35	−2.10 (−5.50 to 1.29)	74.9	0.046	Random effects	0.33	0.99

MRI-PDFF, magnetic resonance imaging proton density fat fraction; CAP, controlled attenuation parameter. Negative values indicate a decrease from baseline to follow-up. Quantitative synthesis included only rows with an available change-score standard deviation; no imputation was performed.

changeability. Third, we synthesized direct head-to-head longitudinal responsiveness data comparing quantitative ultrasound and MRI-PDFF change over time, an outcome not addressed by Wu et al.[8] or, to our knowledge, by other prior reviews in this area. In addition to showing numerical patterns broadly consistent with those previously reported, based on an independent data extraction process, our study presents a critical assessment of the source of threshold values and a preliminary synthesis of longitudinal responsiveness, thereby establishing a key framework for the extent to which these accuracy estimates can be applied to independent clinical populations.

Magnetic resonance imaging proton density fat fraction: diagnostic performance and biological meaning

MRI-PDFF, specificity, DORs, and sample size–weighted descriptive mean AUCs across S ≥ 1/S ≥ 2/S ≥ 3 (Table 2) indicate strong discrimination within the included datasets, but most thresholds were selected in the same cohorts in which performance was evaluated. They therefore describe study-level operating points and should not be used as externally validated clinical thresholds.

The strong performance of MRI-PDFF is supported by its technical characteristics. Chemical shift–encoded MRI corrects for major factors affecting fat quantification, including T1 bias, T2* decay, magnetic-field inhomogeneity, and the multipeak spectral characteristics of fat.[5] Unlike liver biopsy, which evaluates a small tissue sample, MRI-PDFF provides a continuous estimate of fat across a substantially larger liver volume. Tang et al.[14] reported that an MRI-PDFF threshold of 6.4% identified S ≥ 1 with 86% sensitivity and 83% specificity, whereas a threshold of 17.4% identified S ≥ 2 with 64% sensitivity and 96% specificity.[14] These results illustrate that MRI-PDFF maintains strong diagnostic value, whereas its optimal operating point varies with the targeted histological grade.

Nevertheless, MRI-PDFF and histology measure related but different constructs: MRI-PDFF estimates a continuous tissue fat fraction over a large sampling volume, whereas histology categorizes the proportion of steatotic hepatocytes within a small biopsy specimen.[1,5] Consequently, disagreement between the two methods may partly reflect biopsy sampling variability rather than MRI measurement error. Moreover, MRI-PDFF quantifies liver fat but does not directly measure hepatocellular ballooning, lobular inflammation, or fibrosis architecture; assessment of disease activity and fibrosis therefore requires complementary biochemical, elastographic, or histological evaluation.[1]

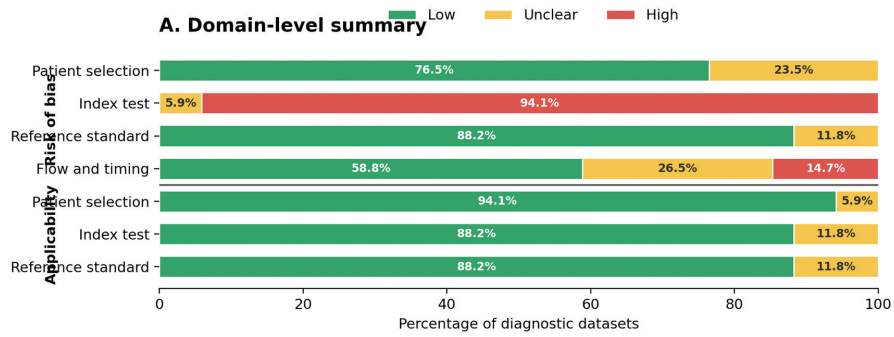


Figure 2. Risk-of-bias and applicability assessment for the 34 diagnostic studies using the revised Quality Assessment of Diagnostic Accuracy Studies, QUADAS-2, tool. The index test domain incorporates explicit assessment of threshold prespecification: study-derived receiver operating characteristic/Youden operating points were rated high-risk, and insufficiently described derivation was rated unclear. Other domain judgments were unchanged. Funnel plots for all diagnostic and longitudinal analyses are consolidated in **Figure 3.** For each diagnostic analysis below, the single embedded panel combines the sensitivity, specificity, and diagnostic odds ratio forest plots (top to bottom); for each longitudinal analysis, the single embedded panel shows the forest plot.

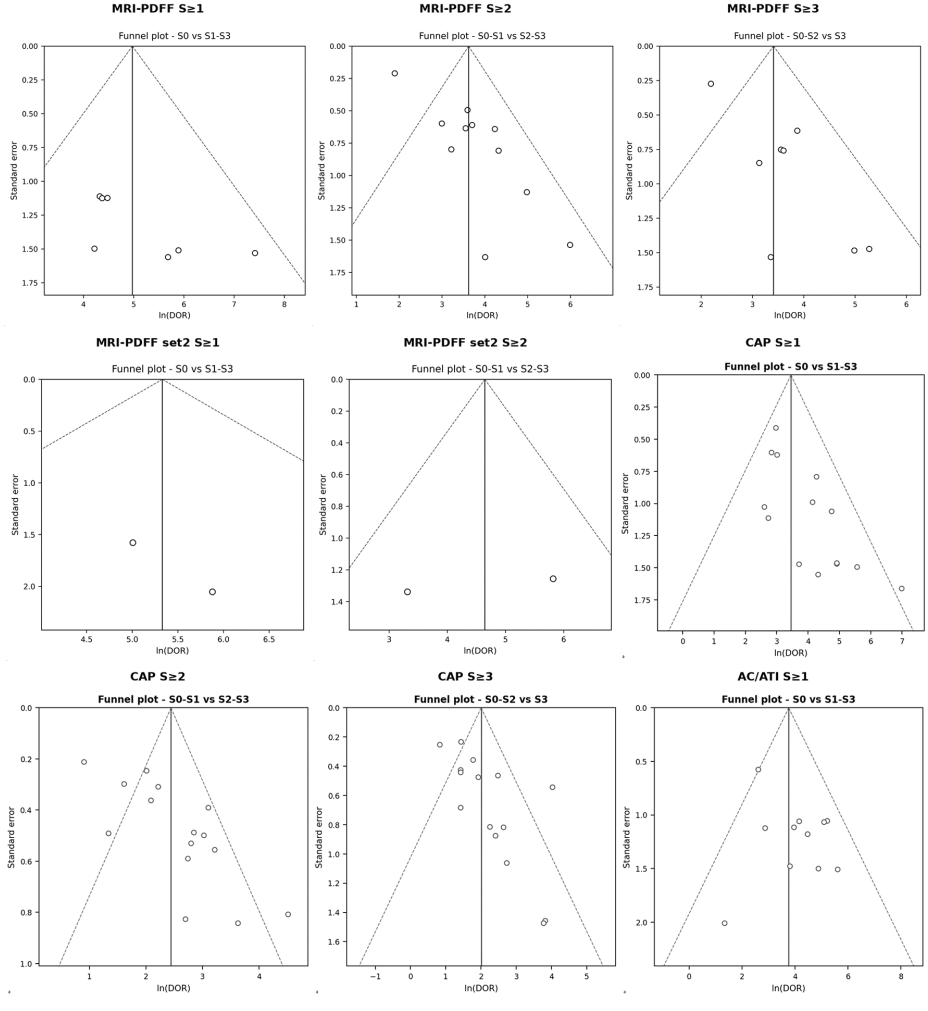


Figure 3. Funnel plots for all diagnostic and longitudinal meta-analyses, grouped by imaging method and histologic threshold. MRI-PDFF, magnetic resonance imaging proton density fat fraction; CAP, controlled attenuation parameter; AC/ATI, attenuation coefficient/attenuation imaging; DOR, diagnostic odds ratio; SE, standard error; UGAP, ultrasound-guided attenuation parameter; UDFF, ultrasound-derived fat fraction.

Controlled attenuation parameter: broad evidence but weaker grade discrimination

The CAP had the largest ultrasound evidence base, but its specificity and DOR declined markedly at higher grades (Table 2), suggesting weaker grade discrimination in the assembled datasets. Between-study variation in disease spectrum, body habitus, probe selection, platform, and cohort-derived cut-offs limits transferability.

The CAP findings are consistent with the major biopsy-referenced studies included in this review, although reported cut-offs and diagnostic performance varied with disease spectrum, body habitus, probe selection, and histological target.[6,7,26,27] At S ≥ 3, pooled specificity was only 64.6%; at the study-selected operating points, approximately one-third of patients without histological S ≥ 3 could therefore be classified as having severe steatosis. Obesity, subcutaneous fat thickness, skin-to-liver capsule distance, probe selection, measurement depth, and acquisition quality may alter CAP attenuation independently of histological fat grade. This combination of patient-, device-, and threshold-related effects limits the transfer of a single cut-off across clinical settings. The CAP remains useful for detecting and following hepatic fat burden, but it should not be used as a stand-alone confirmatory test for S ≥ 3 when precise grade classification would change a clinical or research decision.

Attenuation coefficient/attenuation imaging: Canon attenuation imaging and other attenuation-coefficient implementations

The combined AC/ATI diagnostic estimates (Table 2) should not be interpreted as platform equivalence. Canon ATI is identified separately throughout the study-level extraction, threshold provenance, and QUADAS-2 tables (Supplementary Table 3) because it is not assumed to be interchangeable with other vendors' AC implementations. Only four datasets used Canon ATI in isolation (Dioguardi Burgio et al.,[33] Jang et al.,[37] Zhou et al.,[29] and one Kuroda et al.[3] cohort); for the pooled diagnostic accuracy synthesis reported here, these Canon ATI datasets were combined with other explicitly described AC implementations so that a properly quantified CI could be reported rather than an isolated point estimate without claiming platform equivalence or a formal vendor-specific comparative meta-analysis for Canon ATI alone.

Our findings are consistent with biopsy-referenced ATI studies. Tada et al.[2] reported areas under the ROC curves of 0.85, 0.91, and 0.91 for S ≥ 1, S ≥ 2, and S ≥ 3, respectively. Huang et al.[28] similarly demonstrated high diagnostic performance for ATI in a biopsy-proven cohort, and Dioguardi Burgio et al.[33] reported favorable ATI performance in a prospective comparison with the CAP. Lee et al.[35] and Jang et al.[37] further supported the ability of ATI to identify histologically relevant steatosis categories in multicenter or clinically characterized cohorts. Together,

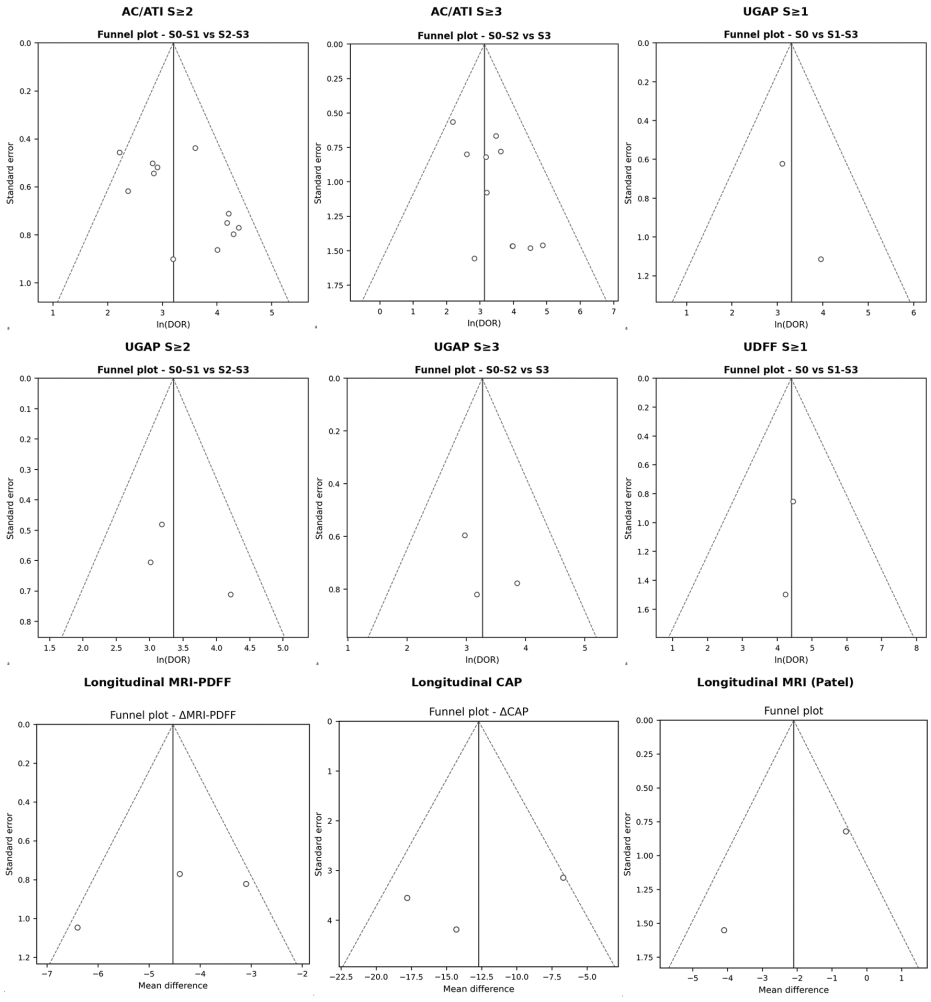


Figure 4. Longitudinal funnel and forest plots for longitudinal magnetic resonance imaging proton density fat fraction (MRI-PDFF) change. **a)** Forest plot for longitudinal MRI-PDFF change. CI, confidence interval.

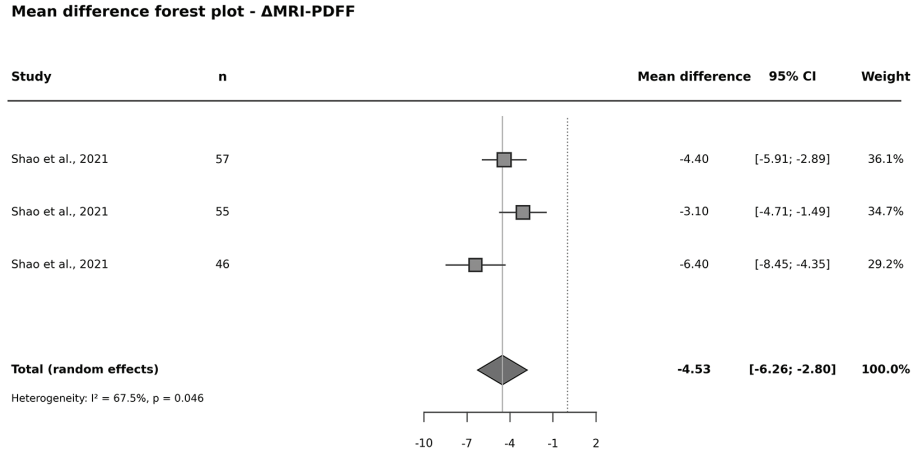


Figure 5. Longitudinal funnel and forest plots for longitudinal controlled attenuation parameter (CAP) change. **a)** Forest plot for longitudinal CAP change.

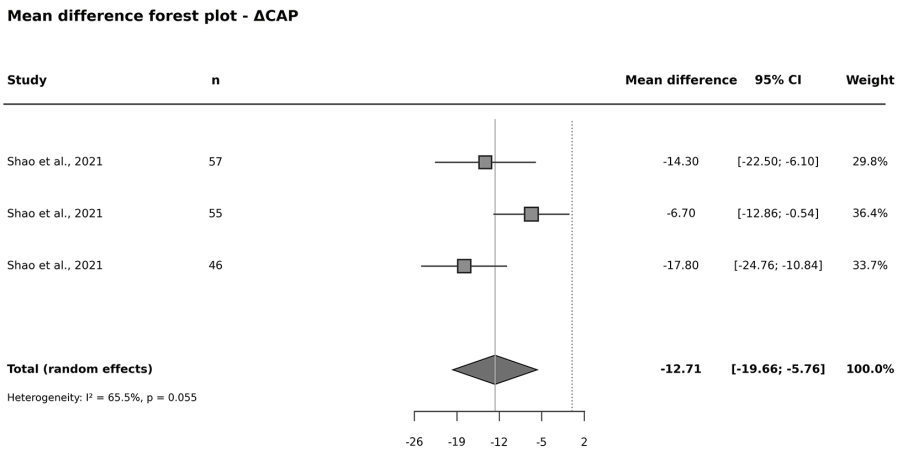


Figure 6. Longitudinal funnel and forest plots for magnetic resonance imaging (MRI) change by histologic response. **a)** Forest plot for MRI change by histologic response.

these studies support the observed increase in ACs with histological fat burden.

Canon ATI and other AC implementations estimate related physical properties but differ in acquisition, calibration, frequency dependence, processing, units, and threshold scale. Platform and manufacturer were therefore retained in the extraction tables; where a primary report did not clearly specify the device model, the manufacturer and model were identified from that study's own Methods Section or device specifications and are documented in Supplementary Table 3, and the discussion avoids transferring a threshold between systems. One Kuroda cohort (n = 125) was evaluated with both Canon ATI and the UGAP; because the same participants were involved, the two modality-specific pooled estimates are not statistically independent, and we did not use them to test whether one modality outperforms the other. A separate Kuroda cohort (n = 202) contributed only to the UGAP analysis.

Ultrasound-guided attenuation parameter and ultrasound-derived fat fraction

The UGAP demonstrated favorable diagnostic performance (Table 2), although only two studies were available at S ≥ 1 and three at S ≥ 2 and S ≥ 3; this small evidence base precludes universal threshold recommendations.

UDFF evidence was limited to two histology-referenced studies (n = 163) at S ≥ 1. Pooled sensitivity was 89.4% (95% CI, 72.8–96.3), specificity was 99.5% (95% CI, 9.4–100.0), and the DOR was 81.67 (95% CI, 19.08–349.57). The extreme width of the specificity CI and substantial heterogeneity (I^2 = 96.0%) indicate marked imprecision; the point estimate should not be interpreted as stable evidence of near-perfect specificity.

Longitudinal responsiveness

Paired longitudinal evidence indicated association rather than interchangeability. An et al.[40] reported absolute and relative change correlations of r = 0.492 ($P < 0.0001$) and r = 0.598 ($P < 0.0001$), with κ = 0.424 ($P = 0.00011$). Shao et al.[11] reported correlations of r = 0.612 and r = 0.524 (both $P < 0.001$), κ = 0.181 ($P = 0.001$), and AUCs of 0.702 ($P = 0.003$) and 0.717 ($P < 0.001$) for identifying MRI-PDFF grade deterioration and improvement. The primary articles did not report CIs for these κ and AUC estimates, and aggregate data were insufficient to reconstruct them. The pooled mean changes (−4.53 percentage points for MRI-PDFF, and −12.71 dB/m for the CAP) are expressed in different units and should not be compared by magnitude.

The clinical interpretation of change is better established for MRI-PDFF. Stine et al.[51] reported that a relative MRI-PDFF reduction of at least 30% was associated with substantially greater odds of histological response; histological response occurred in approximately 51% of MRI-PDFF responders and 14% of non-responders, with a pooled OR of 6.98. This evidence supports the biological relevance of MRI-PDFF change but does not make reduction in liver fat synonymous with MASH resolution or fibrosis regression. In our interpretation, the head-to-head evidence is compatible with MRI-PDFF serving as a more standardized quantitative response measure. This is an interpretation rather than evidence of superiority: only three head-to-head studies were identified, only 158 participants provided analyzable change-score variance, and the available aggregate data do not permit a formal paired comparison of responsiveness between modalities.

The CAP may provide clinically useful serial information when MRI is unavailable, but correlation does not establish agreement, and low-to-moderate κ values show that categorical change can diverge. Hu et al.[31] was not entered into the paired change-score meta-analysis because repeated paired MRI-PDFF change with an extractable variance was unavailable. Future studies should report paired changes, change-score standard deviations, agreement intervals, and externally defined response thresholds for both modalities.

Strengths include histological verification for diagnostic analyses, explicit manufacturer and platform extraction, separate reporting by steatosis threshold, transparent use of Begg's test, and paired longitudinal synthesis when compatible change-score data were available. The study-level tables retain provenance or verification notes so that secondary-source identification is not conflated with primary-source extraction.

Limitations are substantial. Thirty-two of the 34 diagnostic studies selected operating thresholds within the analyzed cohort and 2 were unclear; threshold-related optimism is therefore likely, and the pooled values cannot define external clinical cut-offs. Sensitivity and specificity were pooled separately, which does not model their within-study correlation. A bivariate random-effects or hierarchical summary ROC model was not used because several modality-threshold groups (UGAP, UDFF, and the MRI-PDFF analysis subsets) contained only 2 or 3 studies, in which the between-study correlation of sensitivity and specificity cannot be estimated reliably and hierarchical models frequently fail to converge or yield boundary estimates; for consistency, the same univariate approach was applied to all groups. Hierarchical models should be preferred as the evidence base grows, particularly for the CAP and AC/ATI, for which more studies were available. Consequently, the pooled sensitivity and specificity of a given group do not constitute a jointly estimated summary operating point. Heterogeneity was often substantial, and Begg's test has limited power in small groups. Moreover, AC/ATI combined several non-identical implementations, whereas UGAP and UDFF evidence was sparse. Some datasets contributed to more than one modality analysis, precluding independence for

indirect comparisons. Longitudinal evidence was limited to 3 direct head-to-head studies including 298 participants overall. Only three analyzable change-score rows comprising 158 participants provided the variance data required for quantitative pooling, and missing change-score variances were not imputed.

At study-selected thresholds, MRI-PDFF showed comparatively balanced histology-referenced sensitivity and specificity across steatosis grades. The CAP had the broadest ultrasound evidence but a lower specificity and DOR at higher grades; AC/ATI and UGAP results were promising but platform dependent or based on fewer studies, and UDFF evidence was sparse. Paired longitudinal results indicate an association between CAP and MRI-PDFF change, not interchangeability. Because threshold selection bias, heterogeneity, non-independent datasets, and sparse modality-specific evidence limit inference, these findings should guide future external validation rather than be used as universal cut-offs or as proof of modality superiority.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

Supplemantry Figures Link: https://d2v96fxpocvxx.cloudfront.net/ae2c7771-39a2-40bf-a837-aa639bcdd123/content-images/e977dfd3-9ac4-4282-8739-350ffac7ad71.pdf

Supplemantry Tables Link: https://d2v96fxpocvxx.cloudfront.net/ae2c7771-39a2-40bf-a837-aa639bcdd123/content-images/0c379ca2-71f6-47e5-873b-4904ef8dffa8.pdf

Appendix Table Link: https://d2v96fxpocvxx.cloudfront.net/ae2c7771-39a2-40bf-a837-aa639bcdd123/content-images/f49bd2d8-8197-4a3b-bda1-00aff32336ad.pdf

References

1. European Association for the Study of the Liver (EASL); European Association for the Study of Diabetes (EASD); 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(3):492-542. [Crossref]
2. Tada T, Iijima H, Kobayashi N, et al. Usefulness of attenuation imaging with an ultrasound scanner for the evaluation of hepatic steatosis. *Ultrasound Med Biol.* 2019;45(10):2679-2687. [Crossref]
3. Kuroda H, Abe T, Fujiwara Y, Nagasawa T, Takikawa Y. Diagnostic accuracy of ultrasound-guided attenuation parameter as a noninvasive test for steatosis in non-alcoholic fatty liver disease. *J Med Ultrason (2001).* 2021;48(4):471-480. [Crossref]
4. Labyed Y, Milkowski A. Novel method for ultrasound-derived fat fraction using an integrated phantom. *J Ultrasound Med.* 2020;39(12):2427-2438. [Crossref]
5. Idilman IS, Yildiz AE, Karaosmanoglu AD, Ozmen MN, Akata D, Karcaaltincaba M. Proton density fat fraction: magnetic resonance imaging applications beyond the liver. *Diagn Interv Radiol.* 2022;28(1):83-91. [Crossref]
6. Chan WK, Nik Mustapha NR, Mahadeva S. Controlled attenuation parameter for the detection and quantification of hepatic steatosis in nonalcoholic fatty liver disease. *J Gastroenterol Hepatol.* 2014;29(7):1470-6. [Crossref]
7. Eddowes PJ, Sasso M, Allison M, et al. Accuracy of FibroScan controlled attenuation parameter and liver stiffness measurement in assessing steatosis and fibrosis in patients with nonalcoholic fatty liver disease. *Gastroenterology.* 2019;156(6):1717-1730. [Crossref]
8. Wu S, Pan J, Song M, et al. Performance of magnetic resonance imaging and ultrasound for identifying the different degrees of hepatic steatosis: a systematic review and meta-analysis. *Acad Radiol.* 2025;32(11):6528-6540. [Crossref]
9. 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. [Crossref]
10. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res.* 2018;27(5):1171-1179. [Crossref]
11. Shao CX, Ye J, Dong Z, et al. Steatosis grading consistency between controlled attenuation parameter and MRI-PDFF in monitoring metabolic associated fatty liver disease. *Ther Adv Chronic Dis.* 2021;12:20406223211033119. [Crossref]
12. Kang BK, Yu ES, Lee SS, et al. Hepatic fat quantification: a prospective comparison of magnetic resonance spectroscopy and analysis methods for chemical-shift gradient echo magnetic resonance imaging with histologic assessment as the reference standard. *Invest Radiol.* 2012;47(6):368-375. [Crossref]
13. Idilman IS, Aniktar H, Idilman R, et al. Hepatic steatosis: quantification by proton density fat fraction with MR imaging versus liver biopsy. *Radiology.* 2013;267(3):767-775. [Crossref]
14. Tang A, Desai A, Hamilton G, et al. Accuracy of MR imaging-estimated proton density fat fraction for classification of dichotomized histologic steatosis grades in nonalcoholic fatty liver disease. *Radiology.* 2015;274(2):416-425. [Crossref]
15. Idilman IS, Keskin O, Celik A, et al. A comparison of liver fat content as determined by magnetic resonance imaging-proton density fat fraction and MRS versus liver histology in non-alcoholic fatty liver disease. *Acta Radiol.* 2016;57(3):271-278. [Crossref]
16. Middleton MS, Heba ER, Hooker CA, et al. Agreement between magnetic resonance imaging proton density fat fraction measurements and pathologist-assigned steatosis grades of liver biopsies from adults with nonalcoholic steatohepatitis. *Gastroenterology.* 2017;153(3):753-761. [Crossref]
17. Paige JS, Bernstein GS, Heba E, et al. A pilot comparative study of quantitative ultrasound, conventional ultrasound, and MRI for predicting histology-determined steatosis grade in adult nonalcoholic fatty liver disease. *AJR Am J Roentgenol.* 2017;208(5):W168-W177. [Crossref]
18. Park CC, Nguyen P, Hernandez C, et al. Magnetic resonance elastography vs transient elastography in detection of fibrosis and noninvasive measurement of steatosis in patients with biopsy-proven nonalcoholic fatty liver disease. *Gastroenterology.* 2017;152(3):598-607.e2. [Crossref]
19. Beyer C, Hutton C, Andersson A, et al. Comparison between magnetic resonance and ultrasound-derived indicators of hepatic steatosis in a pooled NAFLD cohort. *PLoS One.* 2021;16(4):e0249491. [Crossref]
20. Kim JW, Lee CH, Yang Z, Kim BH, Lee YS, Kim KA. The spectrum of magnetic resonance imaging proton density fat fraction (MRI-PDFF), magnetic resonance spectroscopy (MRS), and two different histopathologic methods (artificial intelligence *vs.* pathologist) in quantifying hepatic steatosis. *Quant Imaging Med Surg.* 2022;12(11):5251-5262. [Crossref]
21. Nogami A, Yoneda M, Iwaki M, et al. Diagnostic comparison of vibration-controlled transient elastography and MRI techniques in overweight and obese patients with NAFLD. *Sci Rep.* 2022;12(1):21925. [Crossref]
22. Gottfriedova H, Dezortova M, Sedivy P, et al. Comparison of ultrasound to MR and histological methods for liver fat quantification. *Eur J Radiol.* 2025;183:111931. [Crossref] Erratum in: *Eur J Radiol.* 2025;185:112010. [Crossref]

23. Runge JH, Smits LP, Verheij J, et al. MR spectroscopy-derived proton density fat fraction is superior to controlled attenuation parameter for detecting and grading hepatic steatosis. *Radiology.* 2018;286(2):547-556. [Crossref]
24. de Lédinghen V, Wong GL, Vergniol J, et al. Controlled attenuation parameter for the diagnosis of steatosis in non-alcoholic fatty liver disease. *J Gastroenterol Hepatol.* 2016;31(4):848-855. [Crossref]
25. Lee HW, Park SY, Kim SU, et al. Discrimination of nonalcoholic steatohepatitis using transient elastography in patients with nonalcoholic fatty liver disease. *PLoS One.* 2016;11(6):e0157358. [Crossref]
26. Naveau S, Voican CS, Lebrun A, et al. Controlled attenuation parameter for diagnosing steatosis in bariatric surgery candidates with suspected nonalcoholic fatty liver disease. *Eur J Gastroenterol Hepatol.* 2017;29(9):1022-1030. [Crossref]
27. Shalimar, Kumar R, Rout G, et al. Body mass index-based controlled attenuation parameter cut-offs for assessment of hepatic steatosis in non-alcoholic fatty liver disease. *Indian J Gastroenterol.* 2020;39(1):32-41. [Crossref]
28. Huang YL, Bian H, Zhu YL, et al. Quantitative diagnosis of nonalcoholic fatty liver disease with ultrasound attenuation imaging in a biopsy-proven cohort. *Acad Radiol.* 2023;30 Suppl 1:S155-S163. [Crossref]
29. Zhou Y, Nie M, Zhou H, et al. Head-to-head comparison of three different US-based quantitative parameters for hepatic steatosis assessment: a prospective study. *Abdom Radiol (NY).* 2024;49(7):2262-2271. [Crossref]
30. Seo JW, Kim YR, Jang JK, et al. Transient elastography with controlled attenuation parameter versus two-dimensional shear wave elastography with attenuation imaging for the evaluation of hepatic steatosis and fibrosis in NAFLD. *Ultrasonography.* 2023;42(3):421-431. [Crossref]
31. Hu R, Wu B, Wang C, et al. Assessment of transient elastography in diagnosing MAFLD and the early effects of sleeve gastrectomy on MAFLD among the Chinese population. *Int J Surg.* 2024;110(4):2044-2054. [Crossref]
32. Sugimoto K, Moriyasu F, Oshiro H, et al. The role of multiparametric US of the liver for the evaluation of nonalcoholic steatohepatitis. *Radiology.* 2020;296(3):532-540. [Crossref]
33. Dioguardi Burgio M, Castera L, Oufighou M, et al. Prospective comparison of attenuation imaging and controlled attenuation parameter for liver steatosis diagnosis in patients with nonalcoholic fatty liver disease and type 2 diabetes. *Clin Gastroenterol Hepatol.* 2024;22(5):1005-1013.e27. [Crossref]
34. Kuroda H, Fujiwara Y, Abe T, et al. Two-dimensional shear wave elastography and ultrasound-guided attenuation parameter for progressive non-alcoholic steatohepatitis. *PLoS One.* 2021;16(4):e0249493. [Crossref]
35. Lee DH, Cho EJ, Bae JS, et al. Accuracy of two-dimensional shear wave elastography and attenuation imaging for evaluation of patients with nonalcoholic steatohepatitis. *Clin Gastroenterol Hepatol.* 2021;19(4):797-805.e7. [Crossref]
36. Ogino Y, Wakui N, Nagai H, Igarashi Y. The ultrasound-guided attenuation parameter is useful in quantification of hepatic steatosis in non-alcoholic fatty liver disease. *JGH Open.* 2021;5(8):947-952. [Crossref]
37. Jang JK, Lee ES, Seo JW, et al. Two-dimensional shear-wave elastography and US attenuation imaging for nonalcoholic steatohepatitis diagnosis: a cross-sectional, multicenter study. *Radiology.* 2022;305(1):118-126. [Crossref]
38. Cannella R, Agnello F, Porrello G, et al. Performance of ultrasound-guided attenuation parameter and 2D shear wave elastography in patients with metabolic dysfunction-associated steatotic liver disease. *Eur Radiol.* 2025;35(4):2339-2350. [Crossref]
39. Nakamura Y, Hirooka M, Koizumi Y, et al. Diagnostic accuracy of ultrasound-derived fat fraction for the detection and quantification of hepatic steatosis in patients with liver biopsy. *J Med Ultrason (2001).* 2025;52(1):85-94. [Crossref]
40. An Z, Liu Q, Zeng W, et al. Relationship between controlled attenuated parameter and magnetic resonance imaging-proton density fat fraction for evaluating hepatic steatosis in patients with NAFLD. *Hepatol Commun.* 2022;6(8):1975-1986. [Crossref]
41. Wang JH, Ou HY, Yen YH, Chen CH, Lu SN. Usefulness of controlled attenuation parameter in detecting and monitoring hepatic steatosis with MRI-PDFF as reference. *Dig Dis Sci.* 2020;65(5):1512-1519. [Crossref]
42. Agarwal L, Aggarwal S, Shalimar, et al. Bariatric surgery in nonalcoholic fatty liver disease (NAFLD): impact assessment using paired liver biopsy and fibroscan. *Obes Surg.* 2021;31(2):617-626. [Crossref]
43. Nixdorf L, Hartl L, Ströhl S, et al. Rapid improvement of hepatic steatosis and liver stiffness after metabolic/bariatric surgery: a prospective study. *Sci Rep.* 2024;14(1):17558. [Crossref]
44. Hu D, Yang H, Yu G, et al. Ultrasound-derived Fat fraction for pre-operative prediction and monitoring of hepatic steatosis normalization after metabolic surgery in patients with obesity. *Ultrasound Med Biol.* 2026;52(6):1062-1071. [Crossref]
45. Noureddin M, Lam J, Peterson MR, et al. Utility of magnetic resonance imaging versus histology for quantifying changes in liver fat in nonalcoholic fatty liver disease trials. *Hepatology.* 2013;58(6):1930-1940. [Crossref]
46. Loomba R, Lawitz E, Mantry PS, et al. The ASK1 inhibitor selonsertib in patients with nonalcoholic steatohepatitis: a randomized, phase 2 trial. *Hepatology.* 2018;67(2):549-559. [Crossref] Erratum in: *Hepatology.* 2018;67(5):2063. [Crossref]
47. Loomba R, Neuschwander-Tetri BA, Sanyal A, et al. Multicenter validation of association between decline in MRI-PDFF and histologic response in NASH. *Hepatology.* 2020;72(4):1219-1229. [Crossref]
48. Huang DQ, Sharpton SR, Amangurbanova M, Tamaki N, Sirlin CB, Loomba R; NAFLD RESEARCH STUDY GROUP. Clinical utility of combined MRI-PDFF and ALT response in predicting histologic response in nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol.* 2023;21(10):2682-2685.e4. [Crossref]
49. Argo CK, Patrie JT, Lackner C, et al. Effects of n-3 fish oil on metabolic and histological parameters in NASH: a double-blind, randomized, placebo-controlled trial. *J Hepatol.* 2015;62(1):190-197. [Crossref]
50. Patel J, Bettencourt R, Cui J, et al. Association of noninvasive quantitative decline in liver fat content on MRI with histologic response in nonalcoholic steatohepatitis. *Therap Adv Gastroenterol.* 2016;9(5):692-701. [Crossref]
51. Stine JG, Munaganuru N, Barnard A, et al. Change in MRI-PDFF and histologic response in patients with nonalcoholic steatohepatitis: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol.* 2021;19(11):2274-2283.e5. [Crossref]