Supplementary Tables 1–5

MRI-PDFF and Selected Quantitative Ultrasound Methods for Hepatic Steatosis

Supplementary Table 1. Complete database-specific search strategies

Database	Coverage and last search	Search strategy
PubMed/MEDLINE	Inception to July 1, 2026	((NAFLD[Title/Abstract] OR MASLD[Title/Abstract] OR NASH[Title/Abstract] OR "nonalcoholic fatty liver disease"[Title/Abstract] OR "metabolic dysfunction-associated steatotic liver disease"[Title/Abstract] OR "hepatic steatosis"[Title/Abstract]) AND ("MRI-PDFF"[Title/Abstract] OR "proton density fat fraction"[Title/Abstract] OR "chemical shift encoded MRI"[Title/Abstract] OR "controlled attenuation parameter"[Title/Abstract] OR CAP[Title/Abstract] OR "attenuation imaging"[Title/Abstract] OR "attenuation coefficient"[Title/Abstract] OR UGAP[Title/Abstract] OR "ultrasound-guided attenuation parameter"[Title/Abstract] OR UDFF[Title/Abstract] OR "ultrasound-derived fat fraction"[Title/Abstract] OR ibiopsy[Title/Abstract] OR histology[Title/Abstract] OR longitudinal[Title/Abstract] OR monitoring[Title/Abstract] OR responsiveness[Title/Abstract]))
Embase	Inception to July 1, 2026	('nonalcoholic fatty liver disease'/exp OR nafld:ab OR masld:ab OR nash:ti,ab OR 'hepatic steatosis':ti,ab) AND ('proton density fat fraction':ti,ab OR 'mri-pdff':ti,ab OR 'chemical shift encoded mri':ti,ab OR 'controlled attenuation parameter':ti,ab OR 'attenuation imaging':ti,ab OR 'attenuation coefficient':ti,ab OR ugap:ti,ab OR udff:ti,ab OR 'ultrasound-derived fat fraction':ti,ab) AND ('liver biopsy'/exp OR histology:ti,ab OR longitudinal:ti,ab OR monitoring:ti,ab OR responsiveness:ti,ab)
Web of Science Core Collection	Inception to July 1, 2026	(TS=(NAFLD OR MASLD OR NASH OR "nonalcoholic fatty liver disease" OR "metabolic dysfunction-associated steatotic liver disease" OR "hepatic steatosis") AND (MRI-PDFF OR "proton density fat fraction" OR "chemical shift encoded MRI" OR "controlled attenuation parameter" OR "attenuation imaging" OR "attenuation coefficient" OR UGAP OR UDFF OR "ultrasound-derived fat fraction") AND (biopsy OR histology OR longitudinal OR monitoring OR responsiveness))
Cochrane Library	Inception to July 1, 2026	((NAFLD OR MASLD OR NASH OR "hepatic steatosis") AND (MRI-PDFF OR "proton density fat fraction" OR "chemical shift encoded MRI" OR "controlled attenuation parameter" OR "attenuation imaging" OR "attenuation coefficient" OR UGAP OR "ultrasound-guided attenuation parameter" OR UDFF OR "ultrasound-derived fat fraction") AND (biopsy OR histology OR longitudinal OR monitoring OR responsiveness):ti,ab,kw

The search was updated through July 1, 2026. Syntax was adapted to the controlled vocabulary and field structure of each database. No study-design filter was applied electronically; eligibility restrictions were applied during screening. No language restrictions were applied. NAFLD, nonalcoholic fatty liver disease; MASLD, metabolic dysfunction-associated steatotic liver disease; NASH, nonalcoholic steatohepatitis; MRI-PDFF, magnetic resonance imaging-derived proton density fat fraction; CAP, controlled attenuation parameter; UGAP, ultrasound-guided attenuation parameter; UDFF, ultrasound-derived fat fraction.

Supplementary Table 2. COSMIN responsiveness assessment of the longitudinal studies

Study	Construct/comparator	Design and interval	Completeness/technical quality	Statistical approach	Principal concern	Overall
An et al.[40]	CAP change vs. MRI-PDFF change	Prospective treatment cohort; 24 weeks	Doubtful: 95/185 contributed paired longitudinal data	Adequate paired correlations and agreement analyses	Attrition and complete-case selection	Doubtful
Wang et al.[41]	CAP change vs. MRI-PDFF change	Retrospective/prospective follow-up; median 399 days; dispersion NR	Adequate paired imaging dataset	Adequate paired change and correlation analysis	Variable follow-up interval	Adequate
Shao et al.[42]	CAP change vs. MRI-PDFF change	Prospective cohorts; 6 months	Adequate within reported cohorts	Adequate ROC, correlation, and agreement analyses	Heterogeneous intervention cohorts	Adequate
Agarwal et al.[43]	CAP change vs. paired histology	Prospective bariatric cohort; 1 year	Doubtful: selected paired-biopsy sample (n = 58)	Appropriate pre/post and histological comparisons	Selection and attrition	Doubtful
Nisdorf et al.[44]	CAP/VCTE/2D-SWE before and after surgery	Prospective bariatric cohort; 3 months	Doubtful: restricted reliable follow-up subset	Appropriate before/after comparisons	Missing and technically unreliable follow-up measurements	Doubtful
Hu et al.[46]	UDFF change and steatosis normalization	Prospective metabolic-surgery cohort	Doubtful: follow-up completion decreased over time	Appropriate longitudinal/prediction analyses	Incomplete follow-up	Doubtful
Noureddin et al.[45]	MRI-fat change vs. paired histology	Prospective therapeutic-trial datasets	Adequate paired imaging and biopsy	Appropriate paired and association analyses	Heterogeneous trial exposures	Adequate
Loomba et al.[48]	MRI-PDFF change vs. paired histology	Randomized phase II trial; 24 weeks	Adequate paired assessment	Appropriate treatment-response analyses	Small treatment subgroups	Adequate
Loomba et al.[47]	MultiMetter FLINT secondary analysis; 72 weeks	MultiMetter FLINT secondary analysis; 72 weeks	Adequate paired substudy	Appropriate ROC and response analyses	Substudy timing windows	Adequate
Huang et al.[49]	MRI-PDFF/ALT response vs. paired histology	Prospective longitudinal cohort	Adequate paired biopsy and imaging	Standardized response definitions	Variation in treatment exposure and biopsy interval	Adequate
Argo et al.[50]	MRI liver-fat change vs. paired histology	Double-blind randomized trial; 1 year	Doubtful: small completed paired sample	Appropriate within-trial comparisons	Small sample; responsiveness was not the primary measurement property objective	Doubtful

Note. Ratings follow COSMIN Box 10 principles and assess the methodological credibility of imaging responsiveness, not the causal efficacy of the intervention. Overall ratings reflect the lowest relevant item rating (worst score counts principle). Reference numbers correspond to the main manuscript. Abbreviations: 2D-SWE, two-dimensional shear wave elastography; ALT, alanine aminotransferase; CAP, controlled attenuation parameter; COSMIN, COnsensus-based Standards for the selection of health Measurement Instruments; MRI-PDFF, magnetic resonance imaging proton density fat fraction; UDFF, ultrasound-derived fat fraction; VCTE, vibration-controlled transient elastography; NR indicates not reported in the primary study; COSMIN, COnsensus-based Standards for the selection of health Measurement Instruments; CAP, controlled attenuation parameter; MRI-PDFF, magnetic resonance imaging-derived proton density fat fraction; NR, not reported; ROC, receiver operating characteristic; VCTE, vibration-controlled transient elastography; 2D-SWE, two-dimensional shear wave elastography; UDFF, ultrasound-derived fat fraction; ALT, alanine aminotransferase.

Supplementary Table 3. Study-specific histology-referenced cut-offs for quantitative ultrasound methods

Technique	Study	Year	Histologic target	Cut-off	Unit	Manufacturer	System/implementation	Threshold selection
CAP	Beyer 2021	2021	S ≥ 1	268.5	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Beyer 2021	2021	S ≥ 2	308.5	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Beyer 2021	2021	S ≥ 3	337.8	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Chan 2014	2014	S ≥ 2	281	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Chan 2014	2014	S ≥ 3	283	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	de Lédinghen 2016	2016	S ≥ 2	310	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	de Lédinghen 2016	2016	S ≥ 3	311	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Eddowes 2019	2019	S ≥ 1	302	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Eddowes 2019	2019	S ≥ 2	331	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Eddowes 2019	2019	S ≥ 3	337	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Hu 2024	2024	S ≥ 1	271	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Hu 2024	2024	S ≥ 2	292	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Hu 2024	2024	S ≥ 3	301	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Huang 2023	2023	S ≥ 1	258.5	dB/m	Echosens	FibroScan; exact model not reported	Study-specific ROC optimum
CAP	Huang 2023	2023	S ≥ 2	300	dB/m	Echosens	FibroScan; exact model not reported	Study-specific ROC optimum
CAP	Huang 2023	2023	S ≥ 3	315	dB/m	Echosens	FibroScan; exact model not reported	Study-specific ROC optimum
CAP	Lee 2016	2016	S ≥ 1	247	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Lee 2016	2016	S ≥ 2	280	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Lee 2016	2016	S ≥ 3	300	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Naveau 2017	2017	S ≥ 1	298	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Naveau 2017	2017	S ≥ 2	303	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Naveau 2017	2017	S ≥ 3	326	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Nogami 2022	2022	S ≥ 1	263	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Nogami 2022	2022	S ≥ 2	283	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Nogami 2022	2022	S ≥ 3	305	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Park 2017	2017	S ≥ 1	261	dB/m	Echosens	FibroScan 502 Touch; M/XL probes	Study-specific ROC optimum
CAP	Park 2017	2017	S ≥ 2	305	dB/m	Echosens	FibroScan 502 Touch; M/XL probes	Study-specific ROC optimum
CAP	Park 2017	2017	S ≥ 3	312	dB/m	Echosens	FibroScan 502 Touch; M/XL probes	Study-specific ROC optimum
CAP	Runge 2018	2018	S ≥ 1	260	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Runge 2018	2018	S ≥ 2	296	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Runge 2018	2018	S ≥ 3	334	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Seo 2023	2023	S ≥ 1	240	dB/m	Echosens	FibroScan; exact model not reported	Study-specific ROC optimum
CAP	Seo 2023	2023	S ≥ 2	285	dB/m	Echosens	FibroScan; exact model not reported	Study-specific ROC optimum
CAP	Seo 2023	2023	S ≥ 3	298	dB/m	Echosens	FibroScan; exact model not reported	Study-specific ROC optimum
CAP	Shalimar 2020	2020	S ≥ 1	285	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Shalimar 2020	2020	S ≥ 2	331	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Shalimar 2020	2020	S ≥ 3	348	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Shao 2021	2021	S ≥ 1	244	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Shao 2021	2021	S ≥ 2	265	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Shao 2021	2021	S ≥ 3	292	dB/m	Echosens	System model not recorded	Study-specific ROC optimum
CAP	Zhou 2024	2024	S ≥ 1	239	dB/m	Echosens	FibroScan PRO; M/XL probes	Study-specific ROC optimum
CAP	Zhou 2024	2024	S ≥ 2	302	dB/m	Echosens	FibroScan PRO; M/XL probes	Study-specific ROC optimum
CAP	Zhou 2024	2024	S ≥ 3	308	dB/m	Echosens	FibroScan PRO; M/XL probes	Study-specific ROC optimum
Canon ATI	Dioguardi Burgio 2024	2024	S ≥ 1	0.59	dB/cm/MHz	Canon Medical Systems	Aplio i800; iBCX1 1–8 MHz	Study-specific ROC optimum
Canon ATI	Dioguardi Burgio 2024	2024	S ≥ 2	0.72	dB/cm/MHz	Canon Medical Systems	Aplio i800; iBCX1 1–8 MHz	Study-specific ROC optimum
Canon ATI	Huang 2023	2023	S ≥ 1	0.671	dB/cm/MHz	Canon Medical Systems	Aplio platform	Study-specific ROC optimum
Canon ATI	Huang 2023	2023	S ≥ 2	0.726	dB/cm/MHz	Canon Medical Systems	Aplio platform	Study-specific ROC optimum
Canon ATI	Huang 2023	2023	S ≥ 3	0.757	dB/cm/MHz	Canon Medical Systems	Aplio platform	Study-specific ROC optimum
Canon ATI	Jang 2022	2022	S ≥ 1	0.62	dB/cm/MHz	Canon Medical Systems	Aplio i800; PVI-475BX 1–8 MHz	Study-specific ROC optimum
Canon ATI	Jang 2022	2022	S ≥ 2	0.7	dB/cm/MHz	Canon Medical Systems	Aplio i800; PVI-475BX 1–8 MHz	Study-specific ROC optimum
Canon ATI	Jang 2022	2022	S ≥ 3	0.78	dB/cm/MHz	Canon Medical Systems	Aplio i800; PVI-475BX 1–8 MHz	Study-specific ROC optimum
Canon ATI	Kuroda 2021 AC-1	2021	S ≥ 1	0.64	dB/cm/MHz	Canon Medical Systems	ATI measurement reported within comparative study	Study-specific ROC optimum
Canon ATI	Kuroda 2021 AC-1	2021	S ≥ 2	0.71	dB/cm/MHz	Canon Medical Systems	ATI measurement reported within comparative study	Study-specific ROC optimum
Canon ATI	Kuroda 2021 AC-1	2021	S ≥ 3	0.75	dB/cm/MHz	Canon Medical Systems	ATI measurement reported within comparative study	Study-specific ROC optimum
Canon ATI	Lee 2021	2021	S ≥ 1	0.64	dB/cm/MHz	Canon Medical Systems	Aplio i900; PVI-475BX 1–8 MHz	Study-specific ROC optimum
Canon ATI	Lee 2021	2021	S ≥ 2	0.7	dB/cm/MHz	Canon Medical Systems	Aplio i900; PVI-475BX 1–8 MHz	Study-specific ROC optimum
Canon ATI	Lee 2021	2021	S ≥ 3	0.73	dB/cm/MHz	Canon Medical Systems	Aplio i900; PVI-475BX 1–8 MHz	Study-specific ROC optimum
Canon ATI	Seo 2023	2023	S ≥ 1	0.62	dB/cm/MHz	Canon Medical Systems	Aplio platform	Study-specific ROC optimum
Canon ATI	Seo 2023	2023	S ≥ 2	0.7	dB/cm/MHz	Canon Medical Systems	Aplio platform	Study-specific ROC optimum
Canon ATI	Seo 2023	2023	S ≥ 3	0.78	dB/cm/MHz	Canon Medical Systems	Aplio platform	Study-specific ROC optimum
Canon ATI	Sugimoto 2020	2020	S ≥ 1	0.67	dB/cm/MHz	Canon Medical Systems	Aplio i800; PVI-475BX 3.5-MHz convex transducer	Study-specific ROC optimum
Canon ATI	Sugimoto 2020	2020	S ≥ 2	0.72	dB/cm/MHz	Canon Medical Systems	Aplio i800; PVI-475BX 3.5-MHz convex transducer	Study-specific ROC optimum
Canon ATI	Sugimoto 2020	2020	S ≥ 3	0.86	dB/cm/MHz	Canon Medical Systems	Aplio i800; PVI-475BX 3.5-MHz convex transducer	Study-specific ROC optimum
Canon ATI	Tada 2019	2019	S ≥ 1	0.71	dB/cm/MHz	Canon Medical Systems	Aplio platform; exact model not reported in the available article	Threshold derivation insufficiently reported
Canon ATI	Tada 2019	2019	S ≥ 2	0.74	dB/cm/MHz	Canon Medical Systems	Aplio platform; exact model not reported in the available article	Threshold derivation insufficiently reported
Canon ATI	Tada 2019	2019	S ≥ 3	0.74	dB/cm/MHz	Canon Medical Systems	Aplio platform; exact model not reported in the available article	Threshold derivation insufficiently reported
Canon ATI	Zhou 2024	2024	S ≥ 1	0.61	dB/cm/MHz	Canon Medical Systems	Aplio i800; convex transducer 1–8 MHz	Study-specific ROC optimum
Canon ATI	Zhou 2024	2024	S ≥ 2	0.72	dB/cm/MHz	Canon Medical Systems	Aplio i800; convex transducer 1–8 MHz	Study-specific ROC optimum
Canon ATI	Zhou 2024	2024	S ≥ 3	0.77	dB/cm/MHz	Canon Medical Systems	Aplio i800; convex transducer 1–8 MHz	Study-specific ROC optimum
UGAP	Cannella 2025	2025	S ≥ 2	0.75	dB/cm/MHz	GE Healthcare	UGAP platform; radiologist 1 primary	Study-specific ROC optimum
UGAP	Cannella 2025	2025	S ≥ 3	0.75	dB/cm/MHz	GE Healthcare	UGAP platform; radiologist 1 primary	Study-specific ROC optimum
UGAP	Kuroda 2021 AC-2	2021	S ≥ 2	0.654	dB/cm/MHz	GE Healthcare	LOGIQ E10 9 XDclear 2.0; C1-6-D probe, 4.0 MHz	Study-specific ROC optimum
UGAP	Kuroda 2021 AC-2	2021	S ≥ 3	0.691	dB/cm/MHz	GE Healthcare	LOGIQ E9 XDclear 2.0; C1-6-D probe, 4.0 MHz	Study-specific ROC optimum
UGAP	Kuroda 2021 UGAP	2021	S ≥ 1	0.62	dB/cm/MHz	GE Healthcare	LOGIQ E10; C1-6-D courses probe, 4.0 MHz	Study-specific ROC optimum
UGAP	Kuroda 2021 UGAP	2021	S ≥ 2	0.72	dB/cm/MHz	GE Healthcare	LOGIQ E10; C1-6-D courses probe, 4.0 MHz	Study-specific ROC optimum
UGAP	Kuroda 2021 UGAP	2021	S ≥ 3	0.75	dB/cm/MHz	GE Healthcare	LOGIQ E10; C1-6-D courses probe, 4.0 MHz	Study-specific ROC optimum
UGAP	Ogino 2021 UGAP	2021	S ≥ 1	0.6	dB/cm/MHz	GE Healthcare	LOGIQ E9; C1-6 probe	Study-specific ROC optimum
UGAP	Ogino 2021 UGAP	2021	S ≥ 2	0.71	dB/cm/MHz	GE Healthcare	LOGIQ E9; C1-6 probe	Study-specific ROC optimum
UGAP	Ogino 2021 UGAP	2021	S ≥ 3	0.72	dB/cm/MHz	GE Healthcare	LOGIQ E9; C1-6 probe	Study-specific ROC optimum
Investigational QUS-AC	Paige 2017	2017	S ≥ 2	0.809	dB/cm/MHz	Siemens Healthcare	Acuson S3000; 4C1 probe; phantom-referenced offline analysis	Study-derived/cross-validated thresholds
Investigational QUS-AC	Paige 2017	2017	S ≥ 3	0.815	dB/cm/MHz	Siemens Healthcare	Acuson S3000; 4C1 probe; phantom-referenced offline analysis	Study-derived/cross-validated thresholds

Thresholds were study-specific and were not harmonized across platforms. Kuroda et al.[34] (n = 202) and Ogino et al.[36] were classified under UGAP. Paige et al.[37] was retained as investigational QUS-AC. Primary-source verification established that Zhou et al.[38] used Canon ATI (Aplio i800). Paige et al.[37] had no SI group and therefore contributed no S ≥ 1 estimate. AC, attenuation coefficient; ATI, attenuation imaging; CAP, controlled attenuation parameter; QUS, quantitative ultrasound; ROC, receiver operating characteristic; S ≥ 1, any steatosis; S ≥ 2, moderate-or-greater steatosis; S ≥ 3, severe steatosis; UGAP, ultrasound-guided attenuation parameter.

Supplementary Table 4. Study-specific MRI-PDFF thresholds used for histology-referenced steatosis classification

Study	Year	Histologic target	MRI-PDFF cut-off	Unit	Threshold selection/role	2+2 contribution	Primary-source verification note
Kang et al.[12]	2012	S ≥ 1	2.51 (magnitude); 2.94 (complex)	%	Youden-index-derived	Yes	Both reconstruction-specific thresholds were reported; the extraction did not identify which reconstruction generated the retained 2+2 table.
Tang et al.[16]	2015	S ≥ 1	6.9	%	Cohort-derived ROC threshold	Yes	Cohort-derived threshold corresponding to the retained 2+2 performance; distinct from the prespecified 6.4% validation threshold.
Tang et al.[16]	2015	S ≥ 2	16.4	%	Cohort-derived ROC threshold	Yes	Cohort-derived threshold corresponding to the retained 2+2 performance; distinct from the prespecified 17.4% validation threshold.
Tang et al.[16]	2015	S ≥ 3	23.5	%	Cohort-derived ROC threshold	Yes	Cohort-derived threshold corresponding to the retained 2+2 performance; distinct from the prespecified 22.1% validation threshold.
Idilman et al.[13]	2013	S ≥ 2	15.03	%	Maximum sensitivity plus specificity	Yes	Primary full text; cohort-derived ROC threshold.
Idilman et al.[13]	2015	S ≥ 2	10.0	%	Cohort-derived ROC threshold	Yes	Primary full text; threshold linked to the retained operating point.
Middleton et al.[14]	2017	S ≥ 2	16.3	%	Threshold selected at 90% specificity	Yes	Primary full text.
Middleton et al.[14]	2017	S ≥ 3	21.7	%	Threshold selected at 90% specificity	Yes	Primary full text.
Paige et al.[17]	2017	S ≥ 2	13.45	%	Youden-index-derived	Yes	Primary full text; threshold linked to the retained operating point.
Paige et al.[17]	2017	S ≥ 3	16.83	%	Youden-index-derived	Yes	Primary full text; threshold linked to the retained operating point.
Park et al.[18]	2017	S ≥ 1	3.71	%	Youden-index-derived	Yes	Primary full text, Table 3.
Park et al.[18]	2017	S ≥ 2	13.03	%	Youden-index-derived	Yes	Primary full text, Table 3.
Park et al.[18]	2017	S ≥ 3	16.37	%	Youden-index-derived	Yes	Primary full text, Table 3.
Beyer et al.[19]	2021	S ≥ 1	4.4	%	Youden-index-derived	Yes	Primary full text; optimal ROC operating point.
Beyer et al.[19]	2021	S ≥ 2	17.5	%	Youden-index-derived	Yes	Primary full text; optimal ROC operating point.
Beyer et al.[19]	2021	S ≥ 3	22.4	%	Youden-index-derived	Yes	Primary full text; optimal ROC operating point.
Shao et al.[11]	2021	S ≥ 1	5	%	Study-derived ROC boundary	Yes	Primary full text; grade boundaries were 5.0%, 14.6%, and 24.3%.
Shao et al.[11]	2021	S ≥ 2	14.6	%	Study-derived ROC boundary	Yes	Primary full text.
Shao et al.[11]	2021	S ≥ 3	24.3	%	Study-derived ROC boundary	Yes	Primary full text.
Kim et al.[20]	2022	S ≥ 2	10.9	%	Maximum sensitivity plus specificity	Yes	Primary full text, Table 3.
Kim et al.[20]	2022	S ≥ 3	16.2	%	Maximum sensitivity plus specificity	Yes	Primary full text, Table 3.
Nogami et al.[21]	2022	S ≥ 1	4.4	%	Youden-index-derived from patient-level supplementary data	Yes	Exact raw-data cells: TP=146, FP=1, FN=10, TN=6.
Nogami et al.[21]	2022	S ≥ 2	11.7	%	Youden-index-derived from patient-level supplementary data	Yes	Exact raw-data cells: TP=63, FP=7, FN=20, TN=73.
Nogami et al.[21]	2022	S ≥ 3	12.2	%	Youden-index-derived from patient-level supplementary data	Yes	Exact raw-data cells: TP=27, FP=38, FN=2, TN=96.
Gottfriedova et al.[22]	2025	S ≥ 1	5.850	%	Highest Youden index	Yes	Primary full text; VIBEwhole measurement.
Gottfriedova et al.[22]	2025	S ≥ 2	9.75	%	Highest Youden index	Yes	Primary full text; VIBEwhole measurement.

Operating thresholds were verified against the primary reports or, for Nogami et al.,[21] reconstructed directly from the supplied patient-level supplementary dataset. The pooled models synthesize study-specific operating points rather than a universal MRI-PDFF cut-off. Tang et al.[16] reported both prespecified validation thresholds and cohort-derived thresholds; the latter correspond to the 2+2 data retained in this review. Kang et al.[12] reported two reconstruction-specific cut-offs with identical retained classification. FN, false negative; FP, false positive; MRI-PDFF, magnetic resonance imaging proton density fat fraction; ROC, receiver operating characteristic; S ≥ 1, any steatosis; S ≥ 2, moderate-or-greater steatosis; S ≥ 3, severe steatosis; TN, true negative; TP, true positive; VIBE, volumetric interpolated breath-hold examination.

Supplementary Table 5. Assessment of the QUADAS-2 Index Test domain for threshold prespecification

Study/dataset	Modality	Threshold evidence	QUADAS-2 threshold signaling answer	Index test risk	Rationale
Kang et al.[12]	MRI-PDFF	Youden-index-derived	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Idilman et al.[13]	MRI-PDFF	Maximum sensitivity plus specificity	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Tang et al.[16]	MRI-PDFF	Cohort-derived ROC thresholds retained in the synthesis	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Idilman et al.[13]	MRI-PDFF	Cohort-derived ROC threshold	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Middleton et al.[14]	MRI-PDFF	Threshold selected in-cohort at 90% specificity	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Paige et al.[17]	MRI-PDFF/QUS-AC	Study-derived Youden or cross-validated thresholds	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Park et al.[18]	MRI-PDFF/CAP	Youden-index-derived	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Beyer et al.[19]	MRI-PDFF/CAP	Study-specific ROC optima	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Shao et al.[11]	MRI-PDFF/CAP	Study-derived ROC boundaries	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Kim et al.[20]	MRI-PDFF	Maximum sensitivity plus specificity	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Nogami et al.[21]	MRI-PDFF/CAP	Patient-level or study-specific Youden optima	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Gottfriedova et al.[22]	MRI-PDFF	Highest Youden index	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Runge et al.[31]	CAP	Study-specific operating thresholds	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Chan et al.[5]	CAP	Study-specific ROC optimum	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
de Lédinghen et al.[24]	CAP	Study-specific ROC optimum	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Lee et al.[25]	CAP	Study-specific ROC optimum	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Naveau et al.[26]	CAP	Study-specific ROC optimum	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Eddowes et al.[7]	CAP	Study-specific ROC optimum	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Shalimar et al.[27]	CAP	Study-specific ROC optimum	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Huang et al.[28]	Echosens	Study-specific ROC optimum	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Zhou et al.[29]	CAP/Canon ATI	Study-specific ROC optima	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Seo et al.[30]	Echosens	Study-specific ROC optima	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Hu et al.[33]	CAP	Study-specific ROC optimum	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Sugimoto et al.[32]	Canon ATI	Study-specific ROC optimum	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Dioguardi Burgio et al.[35]	Canon ATI	Study-specific ROC optimum	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Kuroda et al.,[2] (n = 125)	Canon ATI	Study-specific ROC optimum	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Lee et al.[25]	Canon ATI	Study-specific ROC optimum	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Ogino et al.[36]	UGAP	Study-specific ROC optimum	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Jang et al.[37]	Canon ATI	Study-specific ROC optimum	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Kuroda et al.[34] (n = 202)	UGAP	Study-specific ROC optimum	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Cannella et al.[39]	UGAP	Study-specific ROC optimum	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Nakamura et al.[38]	UDFF	Study-derived optimal cut-offs	No	High	The operating threshold was selected or tuned using the analyzed cohort and was not externally prespecified.
Tada et al.[1]	Canon ATI	Threshold appeared study-specific, but derivation was not sufficiently explicit	Unclear	Unclear	The available report did not permit a defensible yes/no judgment for threshold prespecification.
Labyed and Milkowski[4]	UDFF	Model development was described, but histology-operating-threshold prespecification was insufficiently explicit	Unclear	Unclear	The available report did not permit a defensible yes/no judgment for threshold prespecification.

QUADAS-2 asks whether the classification threshold was prespecified. A study-derived ROC/Youden threshold was coded as "No" for the signaling question and as high-risk for the Index Test domain, even when blinding was otherwise adequate. This indicates data-dependent threshold selection and possible optimism, not incorporation bias, because histology remained independent of the index test. This table contains the 34 diagnostic studies assessed in the main manuscript. ATI, attenuation imaging; CAP, controlled attenuation parameter; MRI-PDFF, magnetic resonance imaging proton density fat fraction; QUS-AC, quantitative ultrasound attenuation coefficient; QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies 2; ROC, receiver operating characteristic; UDFF, ultrasound-derived fat fraction; UGAP, ultrasound-guided attenuation parameter.