Annals of Gastroenterology (2024) **37,** 579-587

Accuracy of controlled attenuation parameter for liver steatosis in patients at risk for metabolic dysfunction-associated steatotic liver disease using magnetic resonance imaging: a systematic review and meta-analysis

Konstantinos Malandris[a], Anastasia Katsoula[b], Aris Liakos[a,c], Eleni Bekiari[a,c], Thomas Karagiannis[a,c], Eleni Theocharidou[c], Olga Giouleme[b], Emmanouil Sinakos[d], Apostolos Tsapas[a,c,e]

Aristotle University of Thessaloniki, Greece; University of Oxford, UK

Abstract

Background The controlled attenuation parameter (CAP) enables the noninvasive assessment of liver steatosis. We performed a systematic review and meta-analysis to evaluate the diagnostic accuracy of CAP for identifying liver steatosis in patients at risk for metabolic dysfunction-associated steatotic liver disease (MASLD), using magnetic resonance imaging proton density fat fraction (MRI-PDFF) as the reference standard.

Methods We searched Medline, Embase, Cochrane Library and gray literature sources up to March 2024. We defined MASLD as MRI-PDFF ≥5%. We also assessed the accuracy of CAP for identifying patients with MRI-PDFF ≥10%. We calculated pooled sensitivity and specificity estimates using hierarchical random-effects models. We assessed the risk of bias using the Quality Assessment of Diagnostic Accuracy Studies 2 tool, and the certainty in meta-analysis estimates using the Grading of Recommendations Assessment, Development and Evaluation framework.

Results We included 8 studies with 1116 participants. The prevalence of MASLD ranged from 65.2-93.9%. Pooled sensitivity and specificity of CAP for MRI-PDFF ≥5% were 0.84 (95% confidence interval [CI] 0.79-0.88) and 0.77 (95%CI 0.68-0.84), respectively, with an area under the receiver operating characteristic curve (AUROC) of 0.88. The pooled sensitivity and specificity for MRI-PDFF ≥10% were 0.83 (95%CI 0.80-0.87) and 0.72 (95%CI 0.59-0.82), with an AUROC of 0.85. The certainty in our estimates was low to very low because of the high risk of bias, inconsistency and imprecision.

Conclusions CAP has acceptable diagnostic accuracy for both MRI-PDFF ≥5% and MRI-PDFF ≥10%. Adequately powered and rigorously conducted diagnostic accuracy studies are warranted to establish the optimal CAP thresholds.

Keywords Controlled attenuation parameter, magnetic resonance imaging proton density fat fraction, metabolic dysfunction-associated steatotic liver disease, meta-analysis, diagnostic accuracy

Ann Gastroenterol 2024; 37 (5): 579-587

Introduction

The prevalence of metabolic dysfunction-associated steatotic liver disease (MASLD) is greatly increasing: MASLD affects approximately 25-30% of the adult population [1]. Its advanced form, metabolic dysfunction-associated steatohepatitis (MASH), is among the leading causes of hepatocellular carcinoma and liver transplantation in the United States [2,3]. Despite its recognition as a significant healthcare burden, MASLD remains largely underdiagnosed [4,5].

Liver biopsy is considered the reference standard for MASLD diagnosis [1,2]. However, it is an invasive procedure with inherent limitations, including significant intra- and inter-observer variability and sampling error [6]. Moreover,

Conflict of Interest: AT has received research support from Boehringer Ingelheim. EB and ES have received research support from Novo Nordisk. All other authors declare no conflicts of interest

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

Received 1 April 2024; accepted 3 July 2024; published online 19 August 2024

DOI: https://doi.org/10.20524/aog.2024.0910

histologic assessment is impractical for screening at the population level, hampers patient recruitment in clinical trials, while it can potentially lead to life-threatening complications. Therefore, noninvasive biomarkers are needed in both clinical practice and clinical research as alternatives to liver biopsy. Magnetic resonance imaging-derived proton-density-fat fraction (MRI-PDFF) is an accurate and reproducible imaging modality for the assessment of liver steatosis (LS) that strongly correlates with biopsy results [7,8]. Consequently, MRI-PDFF is widely used in MASH clinical trials, either sequentially or as main criterion for patient selection. Nevertheless, its use in the everyday clinical setting is limited because of its high cost and limited availability.

The controlled attenuation parameter (CAP) is an ultrasound modality that allows for the rapid, noninvasive evaluation of LS [9]. However, consensual CAP thresholds for the diagnosis of steatosis in the context of MASLD are lacking [10]. Several trials and meta-analyses have addressed this issue by evaluating the diagnostic accuracy of CAP against liver biopsy [11-14]. While histology remains the gold standard, a shift is evident in emerging studies favoring MRI-PDFF over biopsy as a reliable alternative for steatosis assessment. The accuracy of CAP in identifying various degrees of steatosis, as defined by MRI-PDFF, has not been fully elucidated. We conducted a systematic review and meta-analysis to assess the diagnostic accuracy of CAP for the detection of hepatic steatosis in patients at risk for MASLD, using MRI-PDFF as the reference standard.

Materials and methods

We report our systematic review and meta-analysis, carried out in accordance with the Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) statement (Supplementary Table 1) [15]. Our review is based on a pre-specified protocol registered in PROSPERO (CRD42023464466).

Eligibility criteria

We included studies that assessed the diagnostic accuracy of CAP for the detection of hepatic steatosis in adults with or at risk for MASLD. A study was eligible for inclusion if hepatic steatosis was assessed by means of MRI-PDFF at predefined thresholds (≥5%, ≥10%). Studies assessing liver steatosis in the context of other diseases (e.g., viral hepatitis, autoimmune hepatitis, or polycystic ovary syndrome), or using different MRI-PDFF thresholds, were excluded. MRI thresholds were selected based on recently published guidance for the management of MASLD by the American Association for the Study of Liver Diseases (AASLD) and the MRI-PDFF values used during screening in several MASH clinical trials [1]. For eligible studies with inadequate data to reconstruct 2×2 classification tables, we contacted the corresponding author for relevant information. If no response was received within 15 days, the study was excluded.

Identification and selection of studies

We searched Medline (via PubMed), EMBASE (via Ovid), and Cochrane Library up to July 2023, without date limitations, and updated our search in Medline in March 2024. Our search strategy included free-text and controlled vocabulary terms (Supplementary Tables 2-4). We also searched conference proceedings from relevant scientific meetings from 2016-2023 (Supplementary Table 5), and hand-searched reference lists of pertinent systematic reviews and included studies. All records were imported into literature review software (DistillerSR). Two independent reviewers screened references initially at title and abstract level and subsequently in full text for eligibility. A senior reviewer arbitrated disagreements at any stage.

Data extraction and quality assessment

Pairs of reviewers performed data extraction independently, using a predesigned and pilot-tested form. Data extraction items included trial characteristics, participants' baseline characteristics, true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) values. We assessed the quality and applicability of eligible studies using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS)-2 tool [16]. Details on quality and applicability assessment are presented in the Supplementary Table 6.

Data synthesis

We extracted and reconstructed classification tables for the performance of the index test from eligible studies. We calculated pooled sensitivity, specificity, positive likelihood ratio (LRp), negative likelihood ratio (LRn), and diagnostic odds ratio (DOR) with 95% confidence intervals (CIs) for both reference standard thresholds (≥5%, ≥10%) following the hierarchical random effects model approach [17]. In addition, we constructed summary receiver operating characteristics (sROC) curves with 95% confidence and prediction regions and evaluated the overall performance of CAP using the area under the ROC (AUROC) curve [18].

Heterogeneity is to be expected in diagnostic accuracy meta-analyses [19]. We used the I^2 statistic and the Cochran Q χ^2 test to quantify heterogeneity, defining high

[a]Clinical Research and Evidence-Based Medicine Unit, Aristotle University of Thessaloniki, Greece (Konstantinos Malandris, Aris Liakos, Eleni Bekiari, Thomas Karagiannis, Apostolos Tsapas); [b]Second Propedeutic Medical Department, Aristotle University of Thessaloniki, Greece (Anastasia Katsoula, Olga Giouleme); [c]Second Medical Department, Aristotle University of Thessaloniki, Greece (Aris Liakos, Eleni Bekiari, Thomas Karagiannis, Eleni Theocharidou); [d]Fourth Medical Department, Aristotle University of Thessaloniki, Greece (Emmanouil Sinakos); [e]Harris Manchester College, University of Oxford, UK (Apostolos Tsapas)

heterogeneity as I^2 ≥50% and/or the result of the Cochran Q test being significant (P<0.05) [20]. To address heterogeneity, we performed several sensitivity and subgroup analyses. More specifically, we performed sensitivity analyses that included only studies published in full text, studies that used a Fibroscan device to acquire CAP measurements, and studies with no concerns regarding applicability based on QUADAS-2. In addition, we performed a sensitivity analysis with the exclusion of case-control studies, as this type of study may lead to biased diagnostic accuracy estimates [21]. We conducted subgroup analyses based on the origin of the study (Asia vs. Europe/USA), the power of the MRI scanner (1.5 vs. 3.0 Tesla), the number of quality criteria used to determine a successful CAP examination (≥2 vs. <2 criteria), and CAP positivity thresholds as recommended by the European Association for the Study of the Liver (EASL) (≥275 dB/m) and the AASLD (≥288 dB/m) [1,10]. In terms of quality criteria, we accepted any combination of the following: more than 10 CAP measurements, interquartile range (IQR)/median <30%, and success rate >60% [22]. We conducted *post hoc* subgroup analysis based on the mean body mass index (BMI) of patients included in primary studies (≥30 kg/m^2 and <30 kg/m^2). All statistical analyses were performed using MetaDTA and STATA statistical software [23,24]. We did not assess small study effect bias with funnel plots or statistical tests, as these methods are not recommended in diagnostic test accuracy meta-analyses [25]. All additional analyses were performed for MRI-PDFF ≥5%, which is the threshold used to define MASLD [1].

Grading of evidence

We used the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to assess certainty in our estimates [26-28]. Two reviewers evaluated inconsistency, indirectness, imprecision, publication bias and risk of bias. A senior reviewer arbitrated disagreements. Details on grading of evidence are presented in Supplementary Table 7.

Results

After duplicate removal, we screened 3171 records, from which we selected 10, describing 8 studies with 1116 patients (Supplementary Fig. 1) [29-36]. The baseline characteristics of these studies are presented in Supplementary Table 8. All studies assessed the diagnostic accuracy of CAP against MRI-PDFF with a threshold of 5% [29-36], while 6 of them provided CAP diagnostic accuracy estimates for MRI-PDFF ≥10% [29-31,34-36]. One study was available solely as a conference abstract [29]. Most studies were single-center, following a prospective design (Supplementary Table 9). Three studies had a case-control design [30,33,36]. Seven studies acquired CAP measurements using both the M and XL probe with a Fibroscan device (Echosens, France) [29-35]. One study acquired CAP measurements using the iLivTouch device (Hisky Med, China) [36]. Most studies employed the region-of-interest (ROI) approach to analyze PDFF maps with a varying number of ROIs. The sample size of the included studies ranged from 69-248 participants. The prevalence of MASLD (defined as MRI-PDFF ≥5%) ranged from 65.2-93.9%. Across all participants, 45% were male and 16.6% had type 2 diabetes (T2D). Mean BMI and age were 29.9 kg/m^2 and 49.6 years, respectively. The average mean alanine aminotransferase and aspartate aminotransferase levels across studies were 37.1 U/L and 29.4 U/L, respectively.

Risk of bias and applicability

Most studies were at high risk of bias because of concerns regarding patient selection and the positivity threshold of CAP, which was based on analysis data (Youden index) rather than being pre-specified (Supplementary Fig. 2). One study was susceptible to concerns regarding applicability because it included patients with MASH, derived from clinical trials assessing the efficacy of pharmacologic interventions. For the remaining studies there were no concerns related to applicability (Supplementary Fig. 3).

Diagnostic accuracy

In our main analysis, CAP sensitivity and specificity for MRI-PDFF ≥5% ranged from 0.74-0.91 and from 0.57-0.92, respectively (Fig. 1). Respective values for the diagnosis of MRI-PDFF ≥10% ranged from 0.79-0.87 and from 0.51-0.87 (Fig. 2). The pooled sensitivity and specificity of CAP for the diagnosis of MRI-PDFF ≥5% was 0.84 (95%CI 0.79-0.88, I^2=72.2%) and 0.77 (95%CI 0.68-0.84, I^2=52.6%), yielding an LRp of 3.67 (95%CI 2.55-5.27), LRn of 0.21 (95%CI 0.16-0.30) and a DOR equal to 17.4 (95%CI 9.6-31.6). The pooled sensitivity, specificity, LRp, LRn and DOR for MRI-PDFF ≥10% were 0.83 (95%CI 0.80-0.87, I^2=14.1%), 0.72 (95%CI 0.59-0.82, I^2=86.4%), 2.97 (95%CI 1.94-4.55), 0.23 (95%CI 0.17-0.31), and 12.8 (95%CI 6.4-25.9) respectively. The AUROCs for MRI-PDFF ≥5% and MRI-PDFF ≥10% were 0.88 (95%CI 0.85-0.90) and 0.85 (95%CI 0.82-0.88), respectively (Fig. 3,4).

Subgroup and sensitivity analyses

Results from subgroup analyses are presented in Table 1. Studies conducted in Europe or the USA had pooled sensitivity and specificity of 0.79 and 0.75, respectively, while studies conducted in Asia had pooled sensitivity and specificity both of 0.91. Studies applying ≥2 quality criteria to determine a valid CAP examination resulted in higher specificity estimates (0.85, 95%CI 0.75-0.91) than those using <2 criteria (0.67, 95%CI 0.49-0.81). Similarly, studies with lower BMI produced higher specificity estimates (0.85, 95%CI 0.73-0.93) compared to studies with a mean BMI ≥30 kg/m^2 (0.72, 95%CI

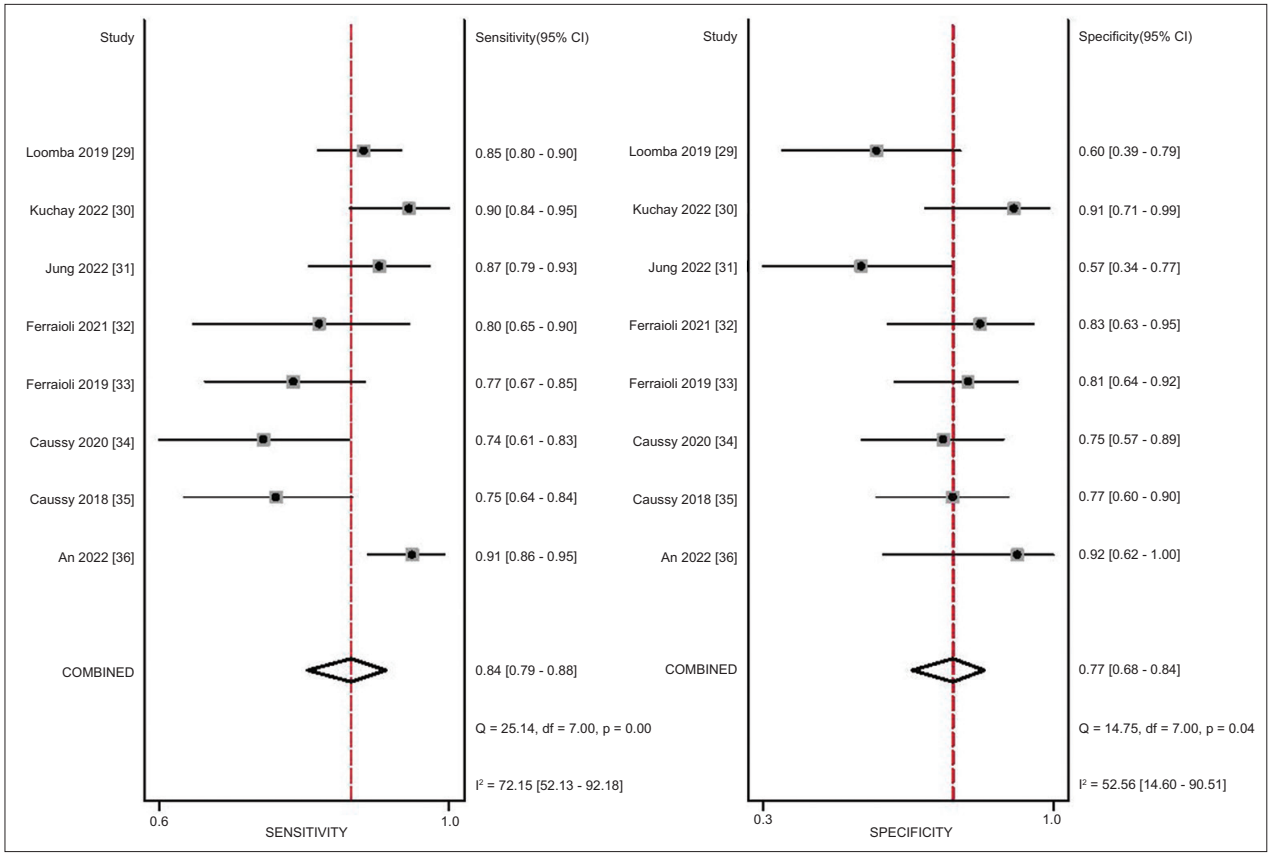


Figure 1 Forest plot of individual study and pooled estimates of sensitivity and specificity of CAP for MRI-PDFF ≥5%
CAP, controlled attenuation parameter; MRI-PDFF, magnetic resonance imaging proton density fat fraction

Table 1 Results from subgroup analyses for MRI-PDFF≥5%

Subgroup analyses	Studies	Patients	Sensitivity (95%CI)	Specificity (95%CI)	LRp (95%CI)	LRn (95%CI)	DOR (95%CI)
Studies conducted in the USA or EU	5	534	0.79 (0.73-0.84)	0.75 (0.66-0.83)	3.18 (2.29-4.40)	0.28 (0.22-0.36)	11.3 (7.2-17.7)
Studies conducted in Asia	2	334	0.91 (0.87-0.94)	0.91 (0.76,0.97)	10.28 (3.49-30.29)	0.10 (0.07-0.15)	100.4 (28.8-349,4)
≥2 quality criteria for CAP measurements	5	645	0.84 (0.77-0.90)	0.85 (0.75-0.91)	5.54 (3.15-9.75)	0.19 (0.12-0.29)	29.6 (11.6-75.5)
<2 quality criterion for CAP measurements	2	223	0.81 (0.69-0.89)	0.67 (0.49-0.81)	2.43 (1.55-3.82)	0.28 (0.18-0.44)	8.7 (4.3-17.4)
BMI <30 kg/m^2	4	576	0.85 (0.76-0.91)	0.85 (0.73-0.93)	5.81 (2.92-11.57)	0.18 (0.10-0.31)	32.7 (10.5-102.2)
BMI ≥30 kg/m^2	3	292	0.81 (0.72-0.87)	0.72 (0.57-0.83)	2.90 (1.86-4.52)	0.27 (0.19-0.38)	10.8 (5.8-20.1)
CAP positivity threshold ≥288 dB/m	4	590	0.81 (0.75-0.87)	0.68 (0.56-0.78)	2.55 (1.84-3.53)	0.28 (0.21-0.36)	9.2 (5.8-14.8)
CAP positivity threshold ≥275 dB/m	5	787	0.84 (0.77-0.89)	0.70 (0.61-0.78)	2.81 (2.12-3.72)	0.23 (0.17-0.33)	12.1 (7.3-20.1)
MRI scanner 3 T	4	539	0.83 (0.74-0.89)	0.73 (0.61-0.83)	3.12 (2.08-4.67)	0.23 (0.15-0.36)	13.6 (6.6-28.1)
MRI scanner 1.5 T	2	192	0.78 (0.70-0.84)	0.82 (0.70-0.90)	4.26 (2.48-7.32)	0.27 (0.19-0.38)	15.8 (7.3-34.3)

BMI, body mass index; PDFF, proton density fat fraction; MRI, magnetic resonance imaging; CI, confidence interval; CAP, controlled attenuation parameter; LRn, negative likelihood ratio; LRp, positive likelihood ratio; DOR, diagnostic odds ratio

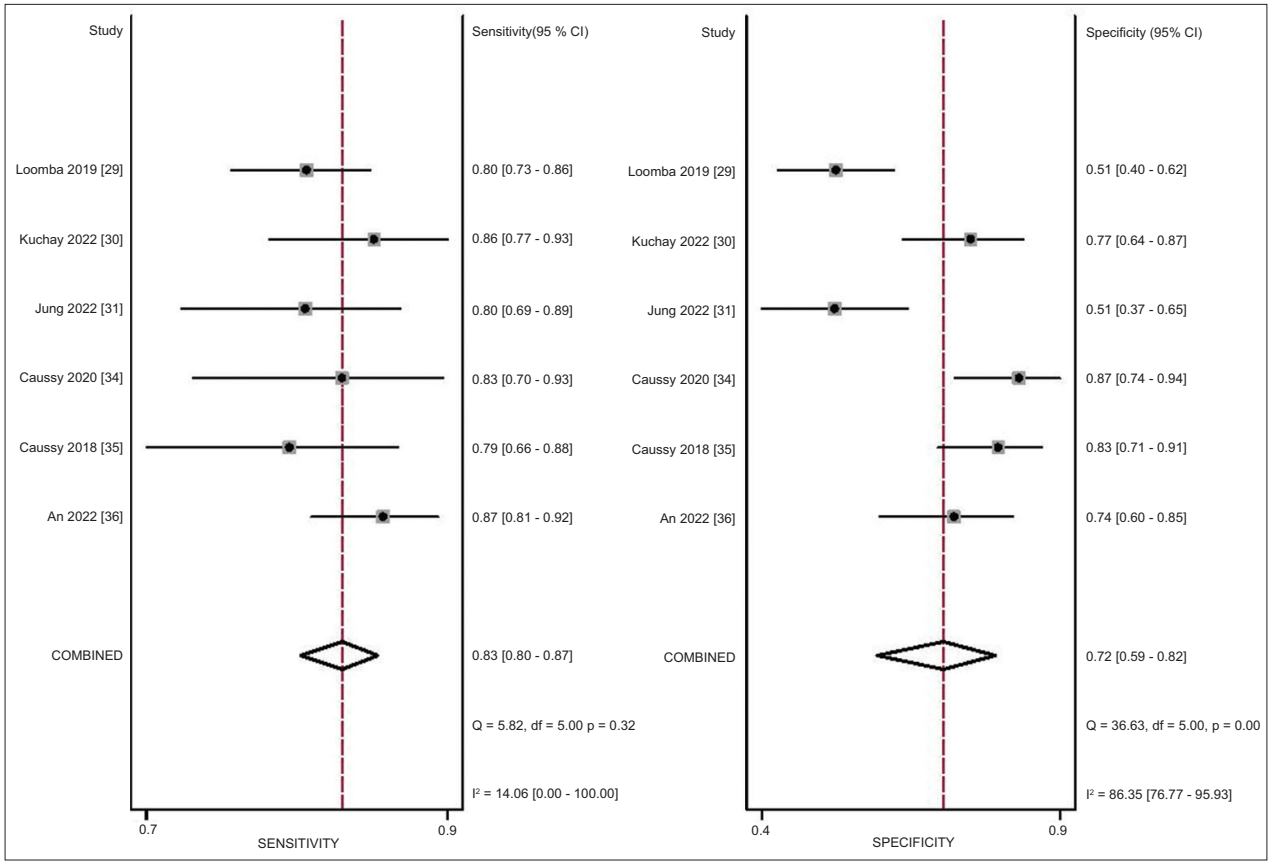


Figure 2 Forest plot of individual study and pooled estimates of sensitivity and specificity of CAP for MRI-PDFF ≥10%
CAP, controlled attenuation parameter; MRI-PDFF, magnetic resonance imaging proton density fat fraction

0.57-0.83). A CAP positivity threshold of ≥288 dB/m (AASLD recommended threshold) resulted in a pooled sensitivity, specificity LRp, LRn and DOR of 0.81, 0.68, 2.55, 0.28 and 9.2 respectively. These estimates were similar for the CAP positivity threshold of ≥275 dB/m recommended by the EASL. Results from sensitivity analyses yielded similar estimates to our main analysis (Supplementary Table 10).

Clinical utility

We explored the clinical utility of CAP in diagnosing MRI-PDFF ≥5% using Fagan nomograms across different prevalence scenarios. In the first scenario, reflecting MASLD prevalence in the general population, we assumed a pre-test probability (prevalence) of 30%. A negative CAP result, with an LRn of 0.2, reduced the post-test probability for MASLD to 8% (Supplementary Fig. 4). In the second scenario, reflecting MASLD prevalence among individuals with obesity and/or T2D, a pre-test probability of 70% was applied. Here, a positive CAP result led to a post-test probability of 90% (Supplementary Fig. 5). Fig. 5 presents the post-test probability for liver steatosis based on pre-test probability and results of CAP testing.

Certainty of evidence

Based on GRADE summaries, the certainty of evidence addressing our research question was low to very low for both MRI-PDFF cutoffs. This was mainly attributed to concerns regarding risk of bias, inconsistency, and imprecision. Details on the evaluation of the certainty of evidence and clarifications on judgments are presented in the Supplementary Tables 11,12.

Discussion

In this systematic review and meta-analysis, we assessed the accuracy of CAP for diagnosing liver steatosis in patients at risk for MASLD, using MRI-PDFF as the reference standard. In addition, we assessed the accuracy of CAP for detecting liver steatosis ≥10%, a threshold commonly employed during screening in MASH clinical trials. CAP demonstrated a good diagnostic accuracy for the detection of both liver steatosis ≥5% (AUROC, 0.88) and liver steatosis ≥10% (AUROC, 0.85). Pooled sensitivity (84% and 83% respectively) and specificity (77% and 72% respectively) estimates were similar for both target conditions. For MRI-PDFF ≥5%, subgroup analyses based on study origin (Asia), BMI (<30 kg/m^2) and a number

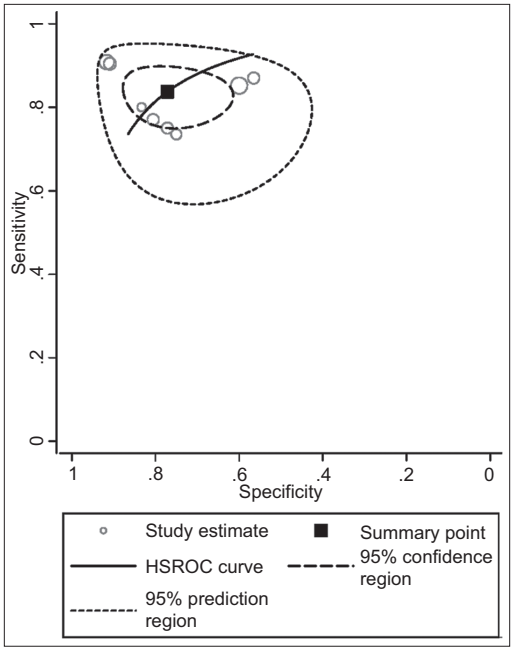


Figure 3 Hierarchical summary receiver operating curve (HSROC) plot of sensitivity versus specificity of CAP for MRI-PDFF ≥5%
Each circle represents a study, with the size proportional to the study size. The curve represents the summary receiver operating characteristic curve for CAP. The square represents the summary estimate of test performance. The bigger-sized dashed outline represents the 95% prediction region and the smaller-sized dashed outline the 95% confidence region
CAP, controlled attenuation parameter; MRI-PDFF, magnetic resonance imaging proton density fat fraction

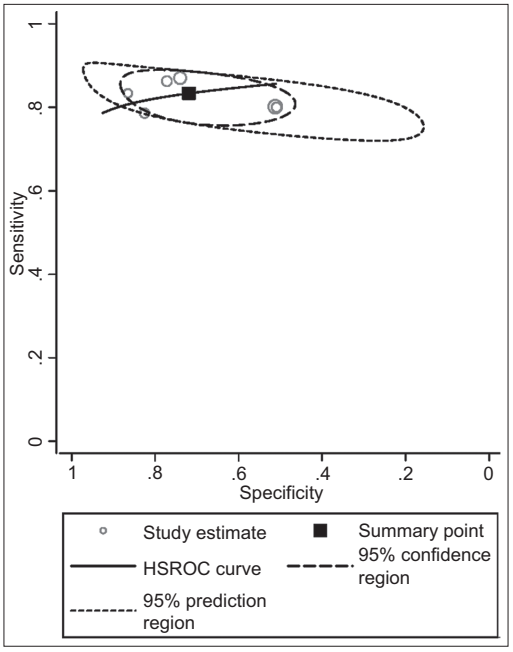


Figure 4 Hierarchical summary receiver operating curve (HSROC) plot of sensitivity vs. specificity of CAP for MRI-PDFF ≥10%
Each circle represents a study, with the size proportional to the study size. The curve represents the summary receiver operating characteristic curve for CAP. The square represents the summary estimate of test performance. The bigger-sized dashed outline represents the 95% prediction region and the smaller-sized dashed outline the 95% confidence region
CAP, controlled attenuation parameter; MRI-PDFF, magnetic resonance imaging proton density fat fraction

of criteria used to determine a successful CAP examination (≥2) resulted in higher diagnostic accuracy estimates.

To our knowledge, this is the first meta-analysis assessing the diagnostic accuracy of CAP in patients at risk for MASLD, using MRI-PDFF as the reference standard. We performed a thorough literature search of major electronic databases and gray literature sources, and contacted authors for additional data in order to produce a comprehensive summary of all the available evidence to date. We performed our analyses using robust methodology, following guidance from the recently published Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy [25]. Moreover, we performed several sensitivity and subgroup analyses to address heterogeneity and assessed the certainty of evidence using the GRADE methodological framework.

Certain limitations should be acknowledged. Despite a thorough literature search, we identified only a small number of studies that met our eligibility criteria. Consequently, our analyses included a small number of studies and participants, and therefore should be interpreted with caution. Most of the studies included in our meta-analysis recruited participants both with and without T2D (proportion of participants with T2D ranging from 14.5-56%). However, data regarding the accuracy of CAP based on the presence of diabetes were sparse, prohibiting subgroup analysis based on T2D status. Furthermore, most of the included studies were at high risk of bias, owing to concerns regarding the positivity thresholds of the index test. More specifically, the positivity threshold for CAP was based on data from analyses (Youden index), rather than being pre-specified. Notably, the studies included in our meta-analysis were conducted prior to the transition from NAFLD to MASLD terminology. Consequently, only 2 studies explicitly outlined the inclusion of metabolic syndrome and/or cardiometabolic comorbidities as criteria [32, 33]. Nevertheless, considering the mean BMI of 29.9 kg/m^2 (range: 28.3-31.7 kg/m^2) across included studies, we are confident that they all align with the MASLD definition, regardless of the initial terminology used. Several studies have examined the overlap between MASLD and NAFLD definitions, indicating high concordance rates of up to 96% [37]. Finally, there was high heterogeneity among studies in both sensitivity and specificity estimates.

Previously published meta-analyses used liver biopsy as the reference standard for MASLD diagnosis [38-40]. In line with our findings, these meta-analyses support a good diagnostic accuracy of CAP for the detection of steatosis ≥S1 (Stage 0 vs. Stage 1-3) and steatosis ≥S2 (Stage 0-1 vs. Stage 2-3), with AUROCs ranging from 0.85-0.96 and from 0.79-0.83, respectively [38-40]. Discrepancies in our results in terms of performance estimates can be primarily attributed to the

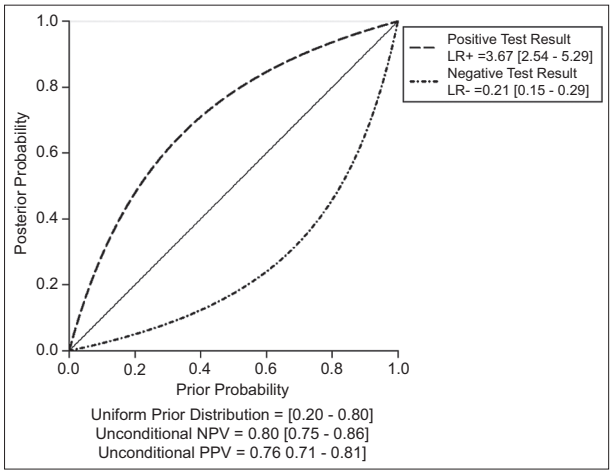


Figure 5 Conditional probability plot for MRI-PDFF ≥5%
The upper dashed curve represents a positive test, and the lower dashed curve a negative test
MRI-PDFF, magnetic resonance imaging proton density fat fraction; LR, likelihood ratio; NPV, negative predictive value; PPV, positive predictive value

different reference standards employed in each meta-analysis (MRI-PDFF vs. liver biopsy). Despite using the same metric (%), histology and PDFF assess steatosis using different approaches. MRI-PDFF offers a continuous measurement directly derived from the properties of the liver, while biopsy results reflect the proportion of hepatocytes with macrovesicular lipid droplets, expressed on an ordinal scale. A recent study evaluated the agreement between the 2 methods, suggesting that above the 5% steatosis threshold, histology estimates tend to exceed PDFF results by up to approximately 3-fold [41]. Furthermore, variations in the prevalence of the target condition and CAP positivity thresholds among primary studies included in each meta-analysis should be taken into consideration. In our study, the median prevalence of MASLD was 76%, while in previous meta-analyses the median prevalence ranged from 90-97%. Meta-epidemiological data suggest an association between higher sensitivity/lower specificity and higher prevalence settings [42]. In our meta-analysis, the average median CAP positivity threshold for MASLD diagnosis was 280 dB/m. In previous meta-analyses, CAP positivity thresholds ranged from 254-273.6 dB/m for steatosis ≥S1 [38-40]. Lower CAP positivity thresholds are likely to favor sensitivity estimates, whereas higher positivity thresholds would enhance specificity results.

In line with our results, previous meta-analyses identified geographic region and BMI as potential sources of heterogeneity [38,39]. Studies conducted in Asia tended to produce higher diagnostic accuracy estimates. Whether or not these findings are related through the lower prevalence of obesity in Asia compared to western countries, and the higher prevalence of lean MASLD in Asia compared to Europe and the USA, is a matter of controversy [43]. Other factors, including dietary and exercise habits, should be taken into consideration before any conclusions are drawn. In addition, there are currently no CAP-specific quality criteria to determine a valid CAP examination. Clinicians commonly apply the already established criteria for liver stiffness measurements. Based on our findings, implementing at least 2 of these criteria (preferably a combination of more than 10 measurements and an IQR/median <30%) can lead to better diagnostic accuracy estimates. Towards this direction, Caussy *et al* suggest that the accuracy of CAP for steatosis detection is more reliable when the IQR of CAP is <30 dB/m [35]. Nevertheless, these CAP-specific quality indicators need further validation before implementation in clinical practice. Our results corroborate findings from a well conducted individual patient meta-analysis assessing CAP using biopsy as reference standard [14]. In contrast to this meta-analysis, we did not identify major differences in the diagnostic performance of CAP between steatosis grades (i.e., ≥5% and ≥10%). Even though there was a trend towards lower AUROCs between the 2 groups, we consider the absolute difference of 5% between the MRI-PDFF thresholds too small to detect major differences.

Based on our findings, the positive and negative LRs of CAP for MRI-PDFF ≥5% were 3.67 and 0.21, respectively. An LRp >10 suggests that a positive test is good at ruling in a diagnosis, whereas an LRn <0.1 is considered clinically meaningful for ruling out a diagnosis [44]. Consequently, the performance of CAP does not endorse its standalone use for ruling in liver steatosis in the context of MASLD. However, in high prevalence settings (i.e., diabetes/obesity/metabolic disease outpatients' clinics) available evidence supports the sequential use of CAP as a triage strategy for ruling out liver steatosis, alongside other noninvasive scores for steatosis (Fatty Liver Index, Hepatic Steatosis Index) and fibrosis (FIB-4 index) [45-47]. CAP positivity thresholds over ≥275 or ≥288 dB/m might be suitable for this purpose [1,10]. An MRI-PDFF value ≥10% is commonly employed for screening participants in MASH clinical trials. Based on our findings, CAP's pooled LRn was 0.23 for MRI-PDFF ≥10%. Therefore, CAP could be used in clinical trials as an initial test to exclude patients with lower degrees of steatosis, thus identifying those patients who are likely to meet other noninvasive inclusion criteria (i.e., MRI-related thresholds) before eventually undergoing liver biopsy. Finally, it should be noted that our results derive mainly from cohorts in tertiary care centers. Currently, there is a paucity of data on the diagnostic performance of CAP using PDFF as reference standard in a primary care setting. To this end, the EPSONIP study (NCT03864510) aims to assess the prevalence and predictors of MASLD in patients with T2D in primary care.

CAP's pooled sensitivity and specificity for MRI-PDFF ≥5% were 0.84 and 0.77, respectively. In a hypothetical cohort of 100 at-risk patients, this implies that 16 individuals would be misclassified as false negatives, indicating the test's failure to identify underlying liver steatosis, and 23 patients would be labeled as false positives, suggesting a misdiagnosis of the target condition. These results highlight the importance of adopting a sequential noninvasive approach to accurately stratify patients at risk for liver steatosis through the implementation of additional risk scores or circulating biomarkers. In the context of a mass screening program, a different diagnostic threshold for CAP with improved sensitivity might be more appropriate,

to effectively rule out the target condition and identify the majority of patients who would require further diagnostic workup.

In conclusion, CAP has acceptable accuracy for the diagnosis of MRI-defined steatosis in patients at risk for MASLD. A pooled sensitivity over 0.80 with an LRn of 0.21 suggests that CAP could be used to effectively rule out steatosis in high-prevalence settings. Further research with adequately powered and rigorously conducted studies is warranted to identify the optimal diagnostic threshold for CAP and establish its place among screening algorithms.

Summary Box

What is already known:

- The controlled attenuation parameter (CAP) enables the noninvasive evaluation of liver steatosis
- Emerging studies favor magnetic resonance imaging proton density fat fraction (MRI-PDFF) over biopsy as a reliable alternative for steatosis assessment
- The accuracy of CAP for liver steatosis in patients at risk for metabolic dysfunction-associated steatotic liver disease (MASLD), using MRI-PDFF as the reference standard, has not been fully elucidated

What the new findings are:

- CAP's pooled sensitivity and specificity for MRI-PDFF ≥5% were 0.84 and 0.77, respectively, with an area under the receiver operating characteristic curve (AUROC) of 0.88
- CAP's pooled sensitivity and specificity for MRI-PDFF ≥10% were 0.83 and 0.72, respectively, with an AUROC of 0.85
- The performance of CAP does not endorse its standalone use for ruling in liver steatosis
- CAP can be used to rule out MRI-defined MASLD

References

1. Rinella ME, Neuschwander-Tetri BA, Siddiqui MS, et al. AASLD Practice Guidance on the clinical assessment and management of nonalcoholic fatty liver disease. *Hepatology* 2023;**77**:1797-1835.
2. Cusi K, Isaacs S, Barb D, et al. American Association of Clinical Endocrinology clinical practice guideline for the diagnosis and management of nonalcoholic fatty liver disease in primary care and endocrinology clinical settings: co-sponsored by the American Association for the Study of Liver Diseases (AASLD). *Endocr Pract* 2022;**28**:528-562.
3. Wong RJ, Aguilar M, Cheung R, et al. Nonalcoholic steatohepatitis is the second leading etiology of liver disease among adults awaiting liver transplantation in the United States. *Gastroenterology* 2015;**148**:547-555.
4. Shiha G, Korenjak M, Eskridge W, et al. Redefining fatty liver disease: an international patient perspective. *Lancet Gastroenterol Hepatol* 2021;**6**:73-79.
5. Alexander M, Loomis AK, Fairburn-Beech J, et al. Real-world data reveal a diagnostic gap in non-alcoholic fatty liver disease. *BMC Med* 2018;**16**:130.
6. Harrison SA, Allen AM, Dubourg J, Noureddin M, Alkhouri N. Challenges and opportunities in NASH drug development. *Nat Med* 2023;**29**:562-573.
7. Caussy C, Reeder SB, Sirlin CB, Loomba R. Noninvasive, quantitative assessment of liver fat by MRI-PDFF as an endpoint in NASH trials. *Hepatology* 2018;**68**:763-772.
8. Middleton MS, Heba ER, Hooker CA, et al; NASH Clinical Research Network. 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**:753-761.
9. Sasso M, Beaugrand M, de Ledinghen V, et al. Controlled attenuation parameter (CAP): a novel VCTE™ guided ultrasonic attenuation measurement for the evaluation of hepatic steatosis: preliminary study and validation in a cohort of patients with chronic liver disease from various causes. *Ultrasound Med Biol* 2010;**36**:1825-1835.
10. European Association for the Study of the Liver. EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis - 2021 update. *J Hepatol* 2021;**75**:659-689.
11. 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**:1717-1730.
12. Siddiqui MS, Vuppalanchi R, Van Natta ML, et al; NASH Clinical Research Network. Vibration-controlled transient elastography to assess fibrosis and steatosis in patients with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol* 2019;**17**:156-163.
13. Karlas T, Petroff D, Sasso M, et al. Individual patient data meta-analysis of controlled attenuation parameter (CAP) technology for assessing steatosis. *J Hepatol* 2017;**66**:1022-1030.
14. Petroff D, Blank V, Newsome PN, et al. Assessment of hepatic steatosis by controlled attenuation parameter using the M and XL probes: an individual patient data meta-analysis. *Lancet Gastroenterol Hepatol* 2021;**6**:185-198.
15. McInnes MDF, Moher D, Thombs BD, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA* 2018;**319**:388-396.
16. Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;**155**:529-536.
17. Macaskill P, Takwoingi Y, Deeks JJ, Gatsonis C. Chapter 9: Understanding meta-analysis. In: Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y (editors). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 2.0 (updated July 2023). Cochrane, 2023. Available from: https://training.cochrane.org/handbook-diagnostic-test-accuracy/current [Accessed 9 July 2024].
18. Takwoingi Y DN, Schiller I, Rücker G, Jones HE, Partlett C, Macaskill P. Chapter 10: Undertaking meta-analysis. In: Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y (editors). Cochrane Handbook for Systematic Reviews of Diagnostic Test, Accuracy. Version 2.0 (updated July 2023). Cochrane. Available from: https://training.cochrane.org/handbook-diagnostic-test-accuracy/current [Accessed 9 July 2024].
19. Lijmer JG, Bossuyt PM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med*

2002;**21**:1525-1537.
20. Cumpston M, Li T, Page MJ, et al. Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. *Cochrane Database Syst Rev* 2019;**10**:ED000142.
21. Reitsma JB, Rutjes AW, Whiting P, et al. Assessing risk of bias and applicability. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy 2023, pp. 169-201.
22. de Lédinghen V, Vergniol J. Transient elastography (FibroScan). *Gastroenterol Clin Biol* 2008;**32**:58-67.
23. Patel A, Cooper N, Freeman S, Sutton A. Graphical enhancements to summary receiver operating characteristic plots to facilitate the analysis and reporting of meta-analysis of diagnostic test accuracy data. *Res Synth Methods* 2021;**12**:34-44.
24. Dwamena BA. midas: a program for meta-analytical integration of diagnostic accuracy studies in Stata. University of Michigan; 2007.
25. Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y (editors). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 2.0 (updated July 2023). Cochrane, 2023. Available from: https://training.cochrane.org/handbook-diagnostic-test-accuracy/current [Accessed 9 July 2024].
26. Schünemann HJ, Oxman AD, Brozek J, et al; GRADE Working Group. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008;**336**:1106-1110.
27. Schünemann HJ, Mustafa RA, Brozek J, et al; GRADE Working Group. GRADE guidelines: 21 part 1. Study design, risk of bias, and indirectness in rating the certainty across a body of evidence for test accuracy. *J Clin Epidemiol* 2020;**122**:129-141.
28. Schünemann HJ, Mustafa RA, Brozek J, et al; GRADE Working Group. GRADE guidelines: 21 part 2. Test accuracy: inconsistency, imprecision, publication bias, and other domains for rating the certainty of evidence and presenting it in evidence profiles and summary of findings tables. *J Clin Epidemiol* 2020;**122**:142-152.
29. Loomba R, Alkhouri N, Patel K, et al. 1727 Validation of cutoffs for controlled attenuation parameter (CAP) with MRI-proton density fat fraction (PDFF) as a reference standard in subjects with nonalcoholic steatohepatitis (NASH) across multiple randomized, controlled trials. *Hepatology* 2019;**70** (Suppl) 1:1040A.
30. Kuchay MS, Choudhary NS, Sharma D, et al. Diagnostic accuracy and optimal cut-off of controlled attenuation parameter for the detection of hepatic steatosis in Indian population. *J Clin Exp Hepatol* 2022;**12**:893-898.
31. Jung J, Han A, Madamba E, et al. Direct comparison of quantitative US versus controlled attenuation parameter for liver fat assessment using MRI proton density fat fraction as the reference standard in patients suspected of having NAFLD. *Radiology* 2022;**304**:75-82.
32. Ferraioli G, Maiocchi L, Savietto G, et al. Performance of the attenuation imaging technology in the detection of liver steatosis. *J Ultrasound Med* 2021;**40**:1325-1332.
33. Ferraioli G, Maiocchi L, Raciti MV, et al. Detection of liver steatosis with a novel ultrasound-based technique: a pilot study using MRI-derived proton density fat fraction as the gold standard. *Clin Transl Gastroenterol* 2019;**10**:e00081.
34. Caussy C, Brissot J, Singh S, et al. Prospective, same-day, direct comparison of controlled attenuation parameter with the M vs the XL probe in patients with nonalcoholic fatty liver disease, using magnetic resonance imaging-proton density fat fraction as the standard. *Clin Gastroenterol Hepatol* 2020;**18**:1842-1850.
35. Caussy C, Alquiraish MH, Nguyen P, et al. Optimal threshold of controlled attenuation parameter with MRI-PDFF as the gold standard for the detection of hepatic steatosis. *Hepatology* 2018;**67**:1348-1359.
36. 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**:1975-1986.
37. Suzuki K, Tamaki N, Kurosaki M, et al. Concordance between metabolic dysfunction-associated steatotic liver disease and nonalcoholic fatty liver disease. *Hepatol Res* 2024;**54**:600-605.
38. Pu K, Wang Y, Bai S, et al. Diagnostic accuracy of controlled attenuation parameter (CAP) as a non-invasive test for steatosis in suspected non-alcoholic fatty liver disease: a systematic review and meta-analysis. *BMC Gastroenterol* 2019;**19**:51.
39. Cao YT, Xiang LL, Qi F, Zhang YJ, Chen Y, Zhou XQ. Accuracy of controlled attenuation parameter (CAP) and liver stiffness measurement (LSM) for assessing steatosis and fibrosis in non-alcoholic fatty liver disease: A systematic review and meta-analysis. *EClinicalMedicine* 2022;**51**:101547.
40. Gu Q, Cen L, Lai J, et al. A meta-analysis on the diagnostic performance of magnetic resonance imaging and transient elastography in nonalcoholic fatty liver disease. *Eur J Clin Invest* 2021;**51**:e13446.
41. Qadri S, Vartiainen E, Lahelma M, et al. Marked difference in liver fat measured by histology vs. magnetic resonance-proton density fat fraction: a meta-analysis. *JHEP Rep* 2023;**6**:100928.
42. Murad MH, Lin L, Chu H, et al. The association of sensitivity and specificity with disease prevalence: analysis of 6909 studies of diagnostic test accuracy. *CMAJ* 2023;**195**:E925-E931.
43. Wong GL-H, Wong VW-S. Non-alcoholic fatty liver disease in Asia: How is it different from the West? *J Gastroenterol Hepatol* 2019;**34**:1267-1268.
44. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet* 2005;**365**:1500-1505.
45. Lee JH, Kim D, Kim HJ, et al. Hepatic steatosis index: a simple screening tool reflecting nonalcoholic fatty liver disease. *Dig Liver Dis* 2010;**42**:503-508.
46. Bedogni G, Bellentani S, Miglioli L, et al. The fatty liver index: a simple and accurate predictor of hepatic steatosis in the general population. *BMC Gastroenterol* 2006;**6**:33.
47. Shah AG, Lydecker A, Murray K, Tetri BN, Contos MJ, Sanyal AJ; Nash Clinical Research Network. Comparison of noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol* 2009;**7**:1104-1112.

Supplementary material

Supplementary Table 1 PRISMA-DTA Checklist

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
		TITLE/ABSTRACT	
Title	1	Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies.	1
Abstract	2	Abstract: See PRISMA-DTA for abstracts.	3
		INTRODUCTION	
Rationale	3	Describe the rationale for the review in the context of what is already known.	4
Clinical role of index test	D1	State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design).	4
Objectives	4	Provide an explicit statement of question (s) being addressed in terms of participants, index test (s), and target condition (s).	4-5
		METHODS	
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	5
Eligibility criteria	6	Specify study characteristics (participants, setting, index test (s), reference standard (s), target condition (s), and study design) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	6, sup info
Search	8	Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated.	sup info
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6, sup info
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	6
Definitions for data extraction	11	Provide definitions used in data extraction and classifications of target condition (s), index test (s), reference standard (s) and other characteristics (e.g. study design, clinical setting).	6
Risk of bias and applicability	12	Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question.	6
Diagnostic accuracy measures	13	State the principal diagnostic accuracy measure (s) reported (e.g. sensitivity, specificity) and state the unit of assessment (e.g. per-patient, per-lesion).	6,7
Synthesis of results	14	Describe methods of handling data, combining results of studies and describing variability between studies. This could include, but is not limited to: a) handling of multiple definitions of target condition. b) handling of multiple thresholds of test positivity, c) handling multiple index test readers, d) handling of indeterminate test results, e) grouping and comparing tests, f) handling of different reference standards	6,7
Meta-analysis	D2	Report the statistical methods used for meta-analyses, if performed.	6,7
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	6,7
		RESULTS	
Study selection	17	Provide numbers of studies screened, assessed for eligibility, included in the review (and included in meta-analysis, if applicable) with reasons for exclusions at each stage, ideally with a flow diagram.	8, sup info
Study characteristics	18	For each included study provide citations and present key characteristics including: a) participant characteristics (presentation, prior testing), b) clinical setting, c) study design, d) target condition definition, e) index test, f) reference standard, g) sample size, h) funding sources	8, sup info

(*Contd...*)

Supplementary Table 1 (*Continued*)

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
Risk of bias and applicability	19	Present evaluation of risk of bias and concerns regarding applicability for each study.	9, sup info
Results of individual studies	20	For each analysis in each study (e.g. unique combination of index test, reference standard, and positivity threshold) report 2×2 data (TP, FP, FN, TN) with estimates of diagnostic accuracy and confidence intervals, ideally with a forest or receiver operator characteristic (ROC) plot.	9, Figure 1, Fig, 2
Synthesis of results	21	Describe test accuracy, including variability; if meta-analysis was done, include results and confidence intervals.	9
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression; analysis of index test: failure rates, proportion of inconclusive results, adverse events).	9,10
		DISCUSSION	
Summary of evidence	24	Summarize the main findings including the strength of evidence.	11
Limitations	25	Discuss limitations from included studies (e.g. risk of bias and concerns regarding applicability) and from the review process (e.g. incomplete retrieval of identified research).	11,12
Conclusions	26	Provide a general interpretation of the results in the context of other evidence. Discuss implications for future research and clinical practice (e.g. the intended use and clinical role of the index test).	12-15
		FUNDING	
Funding	27	For the systematic review, describe the sources of funding and other support and the role of the funders.	2

Supplementary Table 2 Medline (via PubMed)

1. "liver"[tiab] AND ("fatty"[tiab] OR "steatosis"[tiab] OR "steatoses"[tiab])
2. "nash"[All Fields]
3. nafld OR fatty liver OR nash[MeSH Terms]
4. "NAFLD"[tiab]
5. non alcoholic fatty liver disease[MeSH Terms]
6. 1-5/OR
7. "FibroScan"[tiab]
8. "controlled"[All Fields] AND "attenuation"[All Fields] AND "parameter"[All Fields]
9. "controlled attenuation parameter" or "CAP"
10. "controlled attenuation parameter"[tiab] OR "CAP"[tiab]
11. 7-10/OR
12. "diagnos*"[tiab] OR "assess*"[tiab] OR "detect*"[tiab] OR "qualif*"[tiab] OR "discriminat*"[tiab] OR "distin*"[tiab] OR "different*"[tiab] OR "predict*"[tiab] OR "Sensitivity and specificity"[MeSH] OR predict*[tw] OR diagnos*[tw] OR accura*[tw]
13. "Sensitivity and Specificity" [mh]
14. "Likelihood Functions" [mh]
15. "Area Under Curve" [mh]
16. "predictive value of tests" [mh]
17. "Reference Values" [mh]
18. "diagnostic errors" [mh]
19. "Observer Variation" [mh]
20. "false positive*" [tw]
21. "false negative*" [tw]
22. "predictive value*" [tw]
23. "likelihood ratio*" [tw]
24. accurac* [tw]
25. sensitiv* [tw]
26. specificit* [tw]
27. 12-26/OR
28. 6 AND 11 AND 27

Supplementary Table 3 EMBASE (via Ovid)

1. nafld or fatty liver or nash.mp.
2. exp non alcoholic fatty liver disease/
3. exp fatty liver/
4. (liver and (fatty or steatosis or steatoses)).ti, ab.
5. NAFLD.ti, ab.
6. NASH.ti, ab.
7. 1-7/OR
8. 'controlled attenuation parameter':ti, ab, kw
9. cap.mp OR CAP.mp
10. "controlled attenuation parameter".tw. OR CAP.tw.
11. "controlled attenuation parameter" OR CAP
12. controlled.af. AND attenuation.af. AND parameter.af.
13. FibroScan.tw.
14. 8-13/OR
15. exp "Sensitivity and Specificity"/
16. diagnostic error.mp. OR exp "diagnostic errors"/
17. predictive value/
18. exp Area Under Curve/
19. exp Reference Values/
20. exp diagnostic accuracy/OR diagnostic accuracy.mp
21. exp Observer Variation/
22. exp Reproducibility/
23. sensitiv$.mp.
24. specificit$.mp.
25. accurac$.mp.
26. likelihood ratio$.mp.
27. false negative$.mp.
28. false positive$.mp.
29. predictive value$.mp
30. roc curve$.mp. OR exp receiver operating characteristic/
31. "diagnostic odds ratio".mp. OR exp diagnostic value/
32. 'diagnostic test accuracy study'/de
33. 15-32/OR
34. 7 AND 14 AND 33

Supplementary Table 4 Cochrane Library

1. MeSH descriptor: [Non-alcoholic Fatty Liver Disease] explode all trees
2. liver AND (fatty OR steatosis OR steatoses):ti, ab, kw
3. nafld OR fatty liver OR nash: ti, ab, kw
4. NASH OR NAFLD: ti, ab, kw
5. 1-4/OR
6. 'transient elastography':ti, ab, kw OR 'controlled attenuation parameter':ti, ab, kw OR cap: ti, ab, kw OR fibroscan: ti, ab, kw
7. (diagnos* or assess* or detect* or qualif* or discriminat* or distin* or different* or predict*):ti, ab, kw
8. MeSH descriptor: [Sensitivity and Specificity] explode all trees
9. MeSH descriptor: [Area Under Curve] explode all trees
10. ("false positive" OR "false negative" OR "true positive" OR " true negative"):ti, ab, kw
11. 7-10/OR
12. 5 AND 6 AND 11

Supplementary Table 5 Additional gray literature sources searched

American Association for the Study of Liver Diseases (AASLD) (2016-2023)
European Association for the Study of the Liver (EASL) (2016-2023)
United European Gastroenterology Week (UEG) (2016-2023)
American Diabetes Association (ADA) (2016-2023)
European Association for the Study of Diabetes (EASD) (2016-2023)

Supplementary Table 6 Risk of bias and applicability assessment

Participant selection
We judged studies to have low risk of bias if a consecutive recruitment method was used or a random sample was taken from a consecutive series of patients. We also judged studies to have low risk of bias if a case-control design and inappropriate exclusions were avoided. In terms of applicability, a study was judged to have low risk if the participants' spectrum matched our review question.
Index test
We judged studies to have low risk of bias if the CAP results (index test) were interpreted without knowledge of the MRI-PDFF results (reference standard). We also judged studies to have low risk of bias if the CAP positivity threshold was not based on analysis data (i.e., Youden index). In terms of applicability, a study was judged to have low risk if a valid CAP examination was based on at least 10 measurements.
Reference standard
MRI-PDFF is considered to be a highly accurate and reproducible biomarker for steatosis detection that correlates significantly with biopsy results. Consequently, in our systematic review we considered MRI-PDFF to be an adequate method for the detection of the target condition. We judged studies to have low risk of bias if MRI-PDFF was interpreted without knowledge of CAP estimates. We deemed all studies to have low risk in terms of applicability for the reference standard.
Flow and timing
We judged studies to have low risk of bias if the time interval between MRI-PDFF and CAP was ≤3 months. In addition, for a study to have low risk of bias all patients had to undergo an MRI-PDFF, and less than 10% of the patients had to be excluded from the 2×2 tables. Notably, there is no consensus on the acceptable proportion of excluded patients.

Supplementary Table 7 Certainty of evidence assessment

Risk of bias
We ranked this domain as of serious concern. Most included studies were at high risk of bias, mainly due to data-driven positivity thresholds (following the Youden index approach).
Indirectness
We assessed this domain as of not serious concern, based on the applicability judgments of QUADAS-2.
Inconsistency
We ranked this domain as of serious concern. We took into consideration the variance of point estimates of sensitivity and specificity among the included studies, the overlapping of confidence intervals and the size of the 95% prediction regions.
Imprecision
In order to assess imprecision, we defined a minimum acceptable performance level of 0.80 for both sensitivity and specificity. We then checked whether the 95% confidence interval of the pooled sensitivity and specificity estimates crossed the 0.80 level (for which we rated this domain as of serious concern).
Publication bias
Owning to a comprehensive literature search of several databases and gray literature sources we did not downgrade certainty of evidence for publication bias/small study effect concerns.

Supplementary Table 8 Baseline characteristics of included studies

Author, year [ref.]	Study type	Sample size, n	Region	MASLD prevalence (%)	Magnet power (Tesla)	CAP threshold for MRI-PDFF ≥5%	CAP threshold for MRI-PDFF ≥10%	CAP measurement performed with	CAP quality criteria	Probes used	Mean age (years)	Male sex (%)	Diabetes (%)	BMI (kg/m^2)	Liver stiffness (kPa)
An, 2022 [36]	Prospective	197	Asia	93.9	3.0 T	277 dB/m	290.5 dB/m	iLivTouch	>10 measurements, IQR/median <30%	M probe	38.0	169 (85.8)	NR	28.8	7.6
Caussy, 2018 [35]	Prospective	119	USA	70.6	3.0 T	288 dB/m	306 dB/m	FibroScan	>10 measurements, IQR/median <30%, Success rate >60%	M probe XL probe	52.4	49 (41.2)	49 (41.2)	29.9	6.0
Caussy, 2020[#] [34]	Prospective	100	USA	68.0	3.0 T	307 dB/m	322 dB/m	FibroScan	>10 measurements	M probe XL probe	57.5	44 (44.0)	56 (56.0)	30.6	6.6 M probe 5.4 XL probe
Ferraioli, 2019 [33]	Prospective	123	Europe	70.7	1.5 T	258 dB/m	NR	FibroScan	>10 measurements, IQR/median <30%	M probe XL probe	50.4	63 (51.2)	23 (18.8)	29.6	6.1
Ferraioli, 2021* [32]	Prospective	69	Europe	65.2	1.5 T	273 dB/m	NR	FibroScan	>10 measurements, IQR/median <30%,	M probe XL probe	52.8	31 (44.9)	10 (14.5)	30.9	5.1
Jung, 2022 [31]	Prospective	123	USA	81.3	3.0 T	288 dB/m	306 dB/m	FibroScan	>10 measurements	M probe XL probe	52.0	56 (46.0)	47 (38.0)	31.7	NR
Kuchay, 2022 [30]	Prospective	137	Asia	83.9	NR	262 dB/m	295 dB/m	FibroScan	>10 measurements, IQR/median <30%, Success rate >60%	M probe XL probe	44.2	88 (64.2)	NR	28.3	8.4
Loomba, 2019 [29]	Prospective	248	International	90.0	NR	288 dB/m	306 dB/m	FibroScan	NR	M probe XL probe	NR	NR	NR	NR	NR

*MASLD defined as MRI-PDFF>5%

Published in abstract form. Included patients from 3 phase 2 and 3 3 trials including ATLAS and STELLAR-3/-4

[#]All patients underwent measurements with both probes. For our analyses we used the XL probe (the probe with best diagnostic accuracy based on the area under the ROC curve)

NR, Not reported; MASLD, Metabolic dysfunction-associated steatotic liver disease; MRI, magnetic resonance imaging; PDFF, proton density fat fraction; CAP, controlled attenuation parameter

Supplementary Table 9 Additional characteristics of included studies

Author, year [ref.]	Centers	Setting	Recruitment period	MRI-PDFF calculation	Mean ALT (U/L)	Mean AST (U/L)
An, 2022* [36]	single center	hospital	2018-2021	ROIs	69.0	37.0
Caussy, 2018 [35]	single center	NAFLD Research Center	2014-2017	ROIs	48.6	36.4
Caussy, 2020* [34]	single center	NAFLD Research Center	2017-2018	ROIs	45.3	37.4
Ferraioli, 2019 [33]	multicenter	outpatient clinics	2018	NR	26	22.7
Ferraioli, 2021 [32]	multicenter	NR	2018-2020	NR	26.2	23.3
Jung, 2022 [31]	single center	NAFLD Research Center	2015-2019	ROIs	NR	NR
Kuchay, 2022 [30]	single center	Tertiary care facility	2018-2021	ROIs	47.7	35
Loomba, 2019 [29]	multicenter	NASH clinical trials	NR	NR	NR	NR

ALT, alanine aminotransferase; AST, aspartate aminotransferase; NR, Not Reported; NAFLD, nonalcoholic fatty liver disease; MRI, magnetic resonance imaging; PDFF, proton density fat fraction; ROIs, region of interest approach

** ALT, AST values are medians*

Supplementary Table 10 Results from additional sensitivity analyses

Sensitivity analyses	Studies	Patients	Sensitivity (95%CI)	Specificity (95%CI)	LRp (95%CI)	LRn (95%CI)	DOR (95%CI)
Only full text	7	868	0.83 (0.77-0.88)	0.80 (0.71-0.86)	4.10 (2.73-6.03)	0.21 (0.15, 0.30)	19.4 (9.8-38.3)
Excluding case-control	5	659	0.81 (0.75-0.86)	0.71 (0.60-0.80)	2.79 (2.02-3.86)	0.27 (0.21-0.34)	10.4 (6.6-16.2)
Only FibroScan device to acquire CAP measurements	7	919	0.82 (0.77-0.86)	0.75 (0.67-0.82)	3.33 (2.44-4.54)	0.24 (0.18-0.31)	14.0 (8.6-22.8)
No applicability concerns	7	868	0.83 (0.77-0.88)	0.80 (0.71-0.86)	4.10 (2.73-6.03)	0.21 (0.15-0.30)	19.4 (9.8-38.3)

PDFF, proton density fat fraction; MRI, magnetic resonance imaging; CI, confidence interval; CAP, controlled attenuation parameter

Supplementary Table 11 Certainty of evidence for MRI-PDFF ≥5

Sensitivity 0.84 (95%CI 0.79-0.88)			Specificity 0.77 (95%CI 0.68-0.84)					Prevalence 76%	
Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1.000 patients tested	Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 76%	
True positives (patients with NAFLD)	8 studies 1116 patients	cohort & case-control type studies	serious[a]	not serious	serious[b]	serious[c]	none	638 (600 to 669)	⊕◯◯◯ Very low
False negatives (patients incorrectly classified as not having NAFLD)								122 (91 to 160)	
True negatives (patients without NAFLD)	8 studies 1116 patients	cohort & case-control type studies	serious[a]	not serious	serious[b]	serious[c]	none	185 (163 to 202)	⊕◯◯◯ Very low
False positives (patients incorrectly classified as having NAFLD)								55 (38 to 77)	

a. Concerns regarding index test domain in QUADAS-2

b. Variance across studies on point estimates of sensitivity/specificity, degree of 95%CIs overlap between included studies, wide 95% Prediction region

c. Lower boundary of pooled 95%CI for sensitivity and specificity crossing the minimum acceptable performance level of 0.80

MRI, magnetic resonance imaging; PDFF, proton density fat fraction; CI, confidence interval; NAFLD, nonalcoholic fatty liver disease

Supplementary Table 12 Certainty of evidence for MRI-PDFF ≥10%

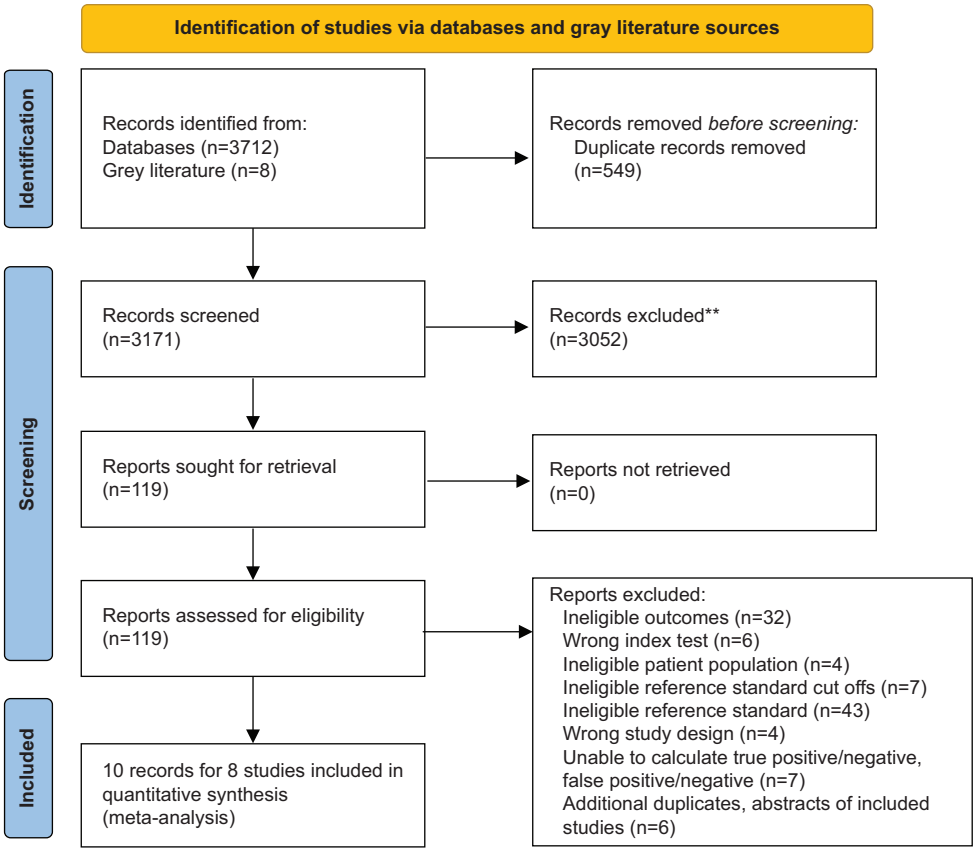
Sensitivity 0.83 (95%CI 0.80-0.87)			Specificity 0.72 (95%CI 0.59-0.82)					Prevalence 58%	
Outcome	№ of studies (№ of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1.000 patients tested	Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 58%	
True positives (patients with MRI-PDFF ≥10%) False negatives (patients incorrectly classified as not having MRI-PDFF ≥10%)	6 studies 924 patients	cohort & case-control type studies	serious[a]	not serious	serious[b]	not serious	none	481 (464 to 505) 99 (75 to 116)	⊕⊕◯◯ Low
True negatives (patients without MRI-PDFF ≥ 10%) False positives (patients incorrectly classified as having MRI-PDFF ≥ 10%)	6 studies 924 patients	cohort & case-control type studies	serious[a]	not serious	serious[b]	serious[C]	none	302 (248 to 344) 118 (76 to 172)	⊕◯◯◯ Very low

a. Concerns regarding index test domain in QUADAS-2

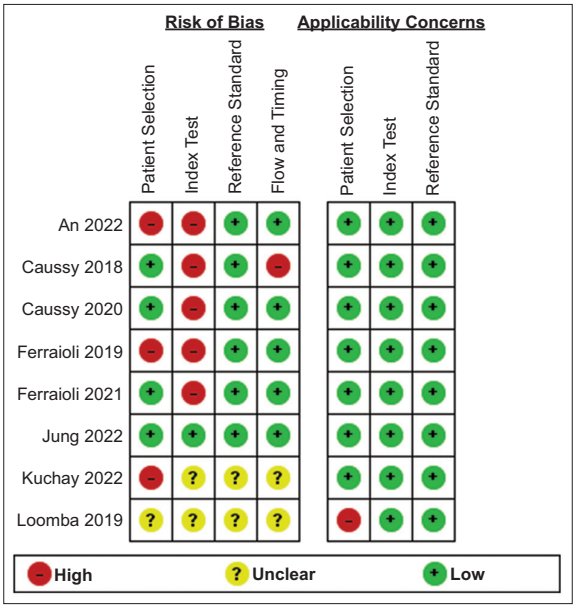
b. Variance across studies on point estimates of sensitivity/specificity, degree of 95%CIs overlap between included studies, wide 95% Prediction region

c. Lower boundary of pooled 95%CI for specificity crossing the minimum acceptable performance level of 0.80

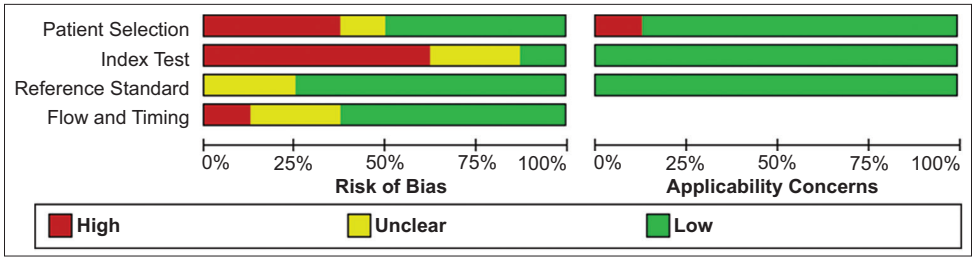
MRI, magnetic resonance imaging; PDFF, proton density fat fraction; CI, confidence interval



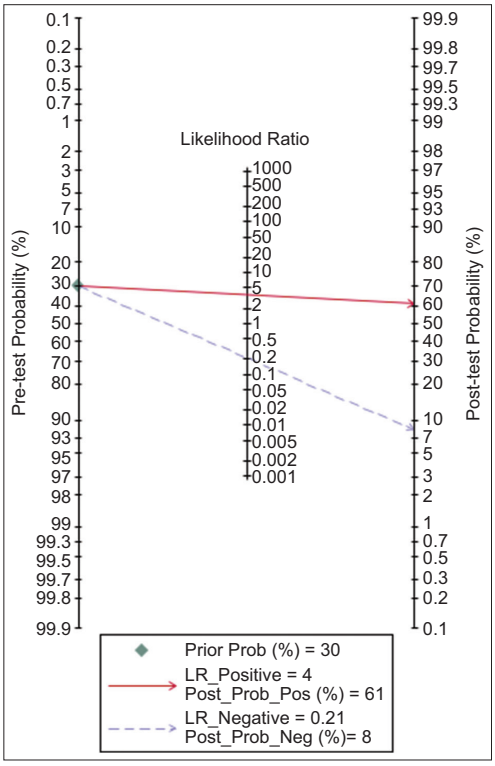
Supplementary Figure 1 PRISMA flow diagram



Supplementary Figure 2 Risk of bias and applicability concerns summary: Judgements about each domain for each included study

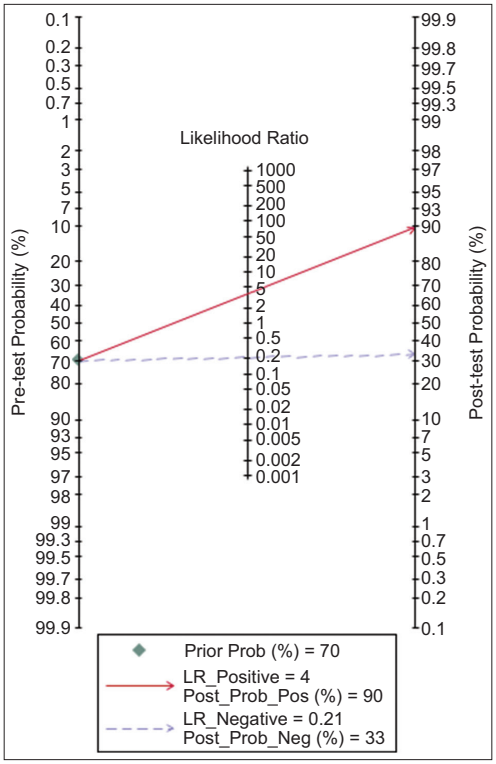


Supplementary Figure 3 Risk of bias and applicability concerns graph. Judgements about each domain presented as percentages across included studies



Supplementary Figure 4 Fagan nomogram for MRI-PDFF ≥5% (pre-test probability 30%)

MRI, magnetic resonance imaging; PDFF, proton density fat fraction; LR, likelihood ratio



Supplementary Figure 5 Fagan nomogram for MRI-PDFF ≥5% (pre-test probability 70%)

MRI, magnetic resonance imaging; PDFF, proton density fat fraction; LR, likelihood ratio