**REVIEW** OPEN ACCESS

# Accuracy of Non-Invasive Imaging Techniques for the Diagnosis of MASH in Patients With MASLD: A Systematic Review

Jennifer Cathcart[1,2] | Rachael Barrett[1] | James S. Bowness[3,4] | Ashis Mukhopadhya[2] | Ruairi Lynch[1] | John F. Dillon[1]

[1]Division of Molecular and Clinical Medicine, University of Dundee, Dundee, UK | [2]Gastroenterology Department, Aberdeen Royal Infirmary, Aberdeen, UK | [3]University College London Hospitals NHS Foundation Trust, London, UK | [4]Department of Targeting Intervention, University College London, London, UK

**Correspondence:** Jennifer Cathcart (090001501@dundee.ac.uk)

## ABSTRACT

**Background and Aims:** Metabolic dysfunction-associated steatotic liver disease (MASLD) is a growing public health problem. The secondary stage in MASLD is steatohepatitis (MASH), the co-existence of steatosis and inflammation, a leading cause of progression to fibrosis and mortality. MASH resolution alone improves survival. Currently, MASH diagnosis is via liver biopsy. This study sought to evaluate the accuracy of imaging-based tests for MASH diagnosis, which offer a non-invasive method of diagnosis.

**Methods:** Eight academic literature databases were searched and references of previous systematic reviews and included papers were checked for additional papers. Liver biopsy was used for reference standard.

**Results:** We report on 69 imaging-based studies. There were 31 studies on MRI, 27 on ultrasound, five on CT, 13 on transient elastography, eight on controlled attenuation parameter (CAP) and two on scintigraphy. The pathological definition of MASH was inconsistent, making it difficult to compare studies. 55/69 studies (79.71%) were deemed high-risk of bias as they had no preset thresholds and no validation. The two largest groups of imaging papers were on MRI and ultrasound. AUROCs were up to 0.93 for MRE, 0.90 for MRI, 1.0 for magnetic resonance spectroscopy (MRS) and 0.94 for ultrasound-based studies.

**Conclusions:** Our study found that the most promising imaging tools are MRI techniques or ultrasound-based scores and confirmed there is potential to utilise these for MASH diagnosis. However, many publications are single studies without independent prospective validation. Without this, there is no clear imaging tool or score currently available that is reliably tested to diagnose MASH.

## 1 | Introduction

The definitions and terminology of fatty liver diseases have recently been revised by a joint consensus of leading experts, endorsed by the American association for study of liver diseases (AASLD) and the European association for study of the liver (EASL). To a spectrum of steatotic liver diseases, of which the disease formally called non-alcoholic fatty liver disease

**Summary**

- There is no clear imaging tool or score currently available to diagnose MASH.
- The most promising imaging tools are MRI techniques or ultrasound-based scores.
- More independent validation studies are needed; this will reduce bias.
- Future work should build on these studies with validation.

(NAFLD) would now most closely approximate metabolic dysfunction-associated steatotic liver disease (MASLD) [1]. MASLD incorporates metabolic dysfunction, the key driver in its pathophysiology. MASLD is estimated to affect 30% of the global population, but with the rise in obesity rates, it is set to increase [2]. Research predicts a 56% rise over the next decade [3]. Thus, it is a public health crisis. Global trends over the past decade have shown a yearly increase in liver transplants for MASLD [4]. In the USA, MASLD is demonstrated to be the fastest growing liver disease over the past five years [5]. It is estimated that MASLD will cost approximately €35 billion annually (ranging between €345 and €1163 per patient) across the UK, Germany, France and Italy and $103 billion ($1613 per patient) in the USA [6]. A German study reported that, in those with advanced disease, mean annual costs were substantially higher, with decompensated cirrhosis costing €22 561, liver transplant patients costing €34 089 and hepatocellular carcinoma (HCC) patients costing €35 910 per patient per year [7]. Dulai et al. found that the stage of liver fibrosis was the strongest predictor for liver-related mortality and the development of other comorbidities [8].

Metabolic dysfunction-associated steatohepatitis (MASH, previously known as NASH), the co-existence of steatosis and inflammation, is a leading cause of progression to fibrosis. The first drug, Rezdiffra (resmetirom), for the treatment of MASH was approved by the FDA in March 2024 [9]. Otherwise, the key management is weight loss strategies. Vilar-Gomez et al. found the highest rates of NASH resolution and fibrosis regression in patients with weight losses ≥ 10% [10]. Adherence to lifestyle modification to achieve this level of weight loss can be difficult for patients. The BRAVES study showed better resolution of NASH in patients after bariatric-metabolic surgery versus lifestyle modification alone [11]. Regression of stage F1 fibrosis is more likely than advanced stages with simple lifestyle changes and treatment of the metabolic comorbidities [12]. Therefore, it is important to correctly diagnose the stage of liver fibrosis and to diagnose at an earlier stage. Thus, diagnosis of MASH, the precursor to fibrosis, is the ideal strategy in the management of MASLD. Lassaily et al. showed that MASH resolution alone improves survival, and 95% fibrosis regression occurs in the presence of MASH resolution [13]. For this reason, we wish to focus on the diagnosis of MASH regardless of fibrosis staging. Currently, the gold standard for diagnosis of MASH is via liver biopsy, which carries significant risks, including mortality [14]. Therefore, a better way to diagnose MASH is needed. There are currently multiple biomarkers that have been developed in response to the need to diagnose MASH, but none have gone on to enter clinical practice.

Previous systematic reviews have focused on MASLD or fibrosis but few on MASH diagnosis. Within these reviews, fewer still have focused on imaging and MASH, with the majority covering biomarkers or clinical scores. We found one previous systematic review focusing on steatohepatitis and imaging by Besutti et al. [15]. This is now 6 years old, and an updated review would be desirable to incorporate further developments in the field. Therefore, we sought to provide a well-informed systematic review to add to the sparse literature on the topic. As there is rising pressure for a non-invasive test for MASH, we felt this review was timely.

## 2 | Materials and Methods

The study was performed after reviewing the SIGN Checklist on systematic reviews [16] and the Cochrane handbook of diagnostic test accuracy methodology [17]. It also followed the PRISMA checklist for diagnostic test accuracy [18] and uploaded a study protocol to Prospero: crd.york.ac.uk/PROSPERO/display_record.php?RecordID=465257.

### 2.1 | Inclusion Criteria

- Patients with a diagnosis of NAFLD/MASLD and a cohort of patients with a diagnosis of MASH/NASH via biopsy.
- Measurable accuracy outcomes: sensitivity, specificity and/or area under the curve of the receiver operating characteristic (AUROC) with comparison to liver biopsy.

### 2.2 | Exclusion Criteria

- Studies on animals or children (Age < 16)
- Studies focusing on fibrosis or steatosis solely
- Abstracts only
- Language other than English

### 2.3 | Reference

Studies reporting liver histology showing MASH via liver biopsy.

### 2.4 | Main Outcome(s)

Primary outcome – Diagnostic accuracy of the test; sensitivity, specificity and/or AUROC or C-statistic.

Secondary outcome – Compare the investigation to the component of the NAS (NAFLD Activity Score) if described.

### 2.5 | Searches

Eight academic databases were searched in Oct 2023 and updated in Nov 2023: Embase, PubMed, Cochrane, CINAHL, ACM digital, clinicaltrials.gov, WHO ICTRP Search Portal and

Open Grey. Advice from an academic librarian was sought to develop and execute a search strategy using a combination of controlled vocabulary and search terms, which was adjusted per database (Supporting Information). Two reviewers (JC and RB) screened the title and abstracts of papers independently after removing duplications using Rayyan Software. Any disagreements were settled between reviewers or adjudicated by a third reviewer (JD). There were no limitations to the publication date. Additionally, references of previous systematic reviews and included papers were checked for additional papers. We also searched trial registries and reported ongoing studies.

## 2.6 | Data Extraction and Risk of Bias and Certainty of Evidence Assessment

One reviewer (JC) extracted demographics from the studies using a predesigned check list. This was cross-checked by a second reviewer (RB). JC and RB evaluated the quality of studies included using the Cochranes quality assessment of diagnostic accuracy studies (QUADAS-C) tool. Any disagreement between the reviewers over the risk of bias was resolved with discussion between reviewers. JC and RB also used the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) evaluation to assess the quality of the papers. JC graded the certainty of evidence using the GRADE (Grading of recommendations assessment, development and evaluation) approach.

## 3 | Results

### 3.1 | Characteristics of Included Studies

A total of 3647 studies plus seven from other sources were found. This is illustrated in the PRISMA flowchart (Supporting Information). After full text review, this was limited to 83 studies. Within the 83 studies, 30 studies reported results either additionally or solely for at-risk MASH (NAS $\geq 4 + \geq$ F2). Sixteen of these studies were reporting on the FibroScan-AST (FAST) score alone. Due to a recent systematic review and meta-analysis on the FAST score in at-risk MASH [19] and as we wish to focus on MASH diagnosis, regardless of fibrosis staging, we have included results for studies that exclusively report on at-risk MASH in the Supporting Information, and here focus analysis on the remaining 69 studies that include results for MASH diagnosis regardless of fibrosis staging, published from 2002 to 2023 [20–88]. The two largest groups of imaging papers were on MRI and ultrasound. There were 31 studies on MRI, 27 on ultrasound, five on CT, 12 on transient elastography, eight on controlled attenuation parameter (CAP) and 2 on scintigraphy. Additionally, there were eight clinical trials. Several studies looked at more than one imaging technique. Table 1 summarises the principles of the different imaging modalities and the theory for MASH diagnosis.

### 3.2 | Demographics

We report data on subject age, gender, body mass index (BMI), race, country of origin and year located in the Supporting Information.

## 3.3 | Histopathological Analysis

Most (30/69 studies) biopsies were percutaneous via ultrasound guidance; the reminder took place intra-operatively, or included both methods, or did not report on the type of biopsy. The pathological definitions of MASH were heterogeneous. The criteria used to diagnose included Brunt criteria, NAS, NASH clinical research network in non-alcoholic hepatitis (NASH CRN), fatty liver inhibition of progression (FLIP) algorithm, steatosis, activity and fibrosis (SAF) score and the Matteoni classification. The most common score was the NAS score, which is calculated by adding scores of three principle components: steatosis (0–3), lobular inflammation (0–3) and hepatocyte ballooning (0–2). The NAS score makes up the NASH CRN score, but NASH CRN also has a separate scoring system for fibrosis, which can be added. The Brunt criteria includes these three parameters but additionally reports on fibrosis. Brunt defines NASH by the presence of fibrosis ($\geq$ grade 1) or necroinflammation ($\geq$ grade 2). It also groups inflammation and ballooning together [89]. The SAF score comprises the grade of activity from A0 to A4, grading both ballooning and lobular inflammation together, and the stage of fibrosis from F0 to F4. The fibrosis staging is like the NASH CRN criteria except it pools the 3 substages (1 a, 1b and 1c) into a single F1 score. Compared with NAS, the SAF activity score equally weights ballooning and lobular inflammation. SAF $>1$ is considered mild disease, $>2$ moderate disease and $>3$ severe disease. The FLIP algorithm diagnoses NASH when at least one of each of the three principle features is present. This contrasts to the NAS score, which can diagnose NASH without one of the components being present [90]. The Matteoni classification defines NAFLD into four types. Types 3 and 4 are what are used in the papers to describe NASH. Type 3; steatosis and hepatocyte ballooning, or type 4; type 3 and Mallory's hyaline or fibrosis [71]. Even within papers using the NAS score, different cut-offs were used for the definition of MASH from NAS $\geq 3$ to NAS $\geq 5$. Several studies also reported using NASH CRN criteria but did not report the specific NAS score used. Other studies had individual definitions of MASH based on the three principle components. Wildman-Tobriner et al.'s study combined 3 trial cohorts, in which two of the cohorts needed fibrosis stages 1–3 to be enrolled in the study [77].

## 3.4 | Risk of Bias

We evaluated the papers using the Cochranes QUADRAS-C toolkit (see Supporting Information). The risk of bias was widespread in the included papers, but no study scored high risk for all four categories: patient selection, index test, reference standard and flow and timing. However, nine studies scored either unclear or high risk within all four categories [22, 32, 41, 44, 65, 77, 83, 85, 86], 11 studies were at risk of selection bias as they only used patients undergoing bariatric surgery [21, 22, 30, 32, 38, 39, 49, 57, 65, 67, 86] and one study excluded all those with BMI $> 40$ or seen in the obesity clinic [72]. One study only included patients with type 2 diabetes [83], one study excluded all patients with type 2 diabetes on insulin [25] and a further study excluded those on glucose-lowering tablets [80]. Three studies included clinical trial patients [40, 56, 77]; one of these excluded any patient with uncontrolled diabetes [77]. Whilst all studies included

**TABLE 1** | Summary of underlying principles of imaging methods and theory in diagnosing MASH.

| | Underlying principles | Theory in diagnosing MASH |
|---|---|---|
| *MRI* | | |
| MRE | Generates mechanical shear waves by passing low-frequency vibrations through the liver delivered via a pad placed over the patient's liver. The wavelength reflects stiffness of the tissue | Detects increased stiffness of liver secondary to developing fibrosis and inflammation related to MASH |
| MRI-PDFF | Compares the fat and water components of an MRI scan, and expresses as a percentage | Detects increased steatosis, which could be a marker of increased risk of MASH. Positive correlations between change in MRI-PDFF and steatosis grade have been demonstrated [40, 77] |
| MpMRI | Combines different MRI techniques into a single scan. T1 mapping detects increases in extracellular fluid. Hepatic iron can be a confounding factor. Therefore, iron is calculated with parallel acquisition of T2. Iron-corrected T1 relaxation maps (cT1) take out this confounding factor | • cT1 or T1 mapping correlates with fibrosis and inflammation. Inflammation creates increases in extracellular fluid detected by the scan [64]<br>• Liver*MultiScan Software (*Perspectum, Oxford, UK) postprocesses non-contrast MpMRI imaging scans to simultaneously quantify steatosis (via MRI-PDFF), fibrosis and inflammation (via cT1) |
| GA-MRI | Gadoxetic acid is a hepatobiliary contrast agent that is taken up by the hepatocytes, allowing more detailed interpretation of liver anatomy | The hepatobiliary phase T1 relaxation time potentially differentiates MASH from steatosis. Amorim et al. hypothesise that expression of the multidrug resistance protein transporter is increased in patients with MASH, causing an increase in GA excretion with a consequent decrease in signal intensity [24]. Bastati et al. hypothesise the reduction in signal intensity is due to liver injuries reducing hepatosinusoidal space and thus GA uptake [26] |
| MRI | MRI creates a magnetic field, which forces protons in the body to align with that field. The protons are then stimulated by a radiofrequency current. When the radiofrequency field is turned off, the MRI sensors detect the energy released as the protons realign with the magnetic field. The difference between various types of tissues is based on these magnetic properties | Different methods<br>• Gallego-Duran et al.: Utilise three MRI variables from three different MRI protocols and call then NASHMRI; they found them to be independently associated with MASH [36]<br>• Leporq et al.: Magnetic susceptibility; the magnetic susceptibility decreased in MASH. Felt to be related to diamagnetic material changes in fibrosis and inflammation [55] |
| IVIM | Based on the random movement of individual water protons, which results in molecular diffusion, called Brownian motion. The mean square distance travelled by a molecule is proportional to time and diffusion coefficient (D). The scan also measures the perfusion component (D*) and the vascular volume fraction (f) | Accumulation of fat, inflammation and hepatocellular ballooning, present in MASH, reduce sinusoidal space and cause restrictive barriers detected as reduced diffusion (D). D* and f indicate change that occurs in the liver blood flow because of distortion of the microcirculatory anatomy and the compression of the sinusoidal space related to inflammation. Additionally, fibrosis can also influence D, D* and f [83] |

TABLE 1 | (Continued)

| | Underlying principles | Theory in diagnosing MASH |
|---|---|---|
| MRS | The distribution of electrons within an atom causes nuclei within different molecules to have a slightly different magnetic field. This results in slightly different resonant frequencies, which creates an image from the different signals. This allows the presence and concentration of various metabolites to be determined from tissues | Inflammation related to MASH causes different metabolic changes, which can be detected by MRS. Kim et al. use $^1$H-MRS (proton MRS) with long echo time; among other metabolites, they looked at Ala peaks, which are shown in other studies to be specific biomarkers for liver injury and indicative of liver enzymes raised by inflammation [48]. Abrigo et al. used phosphorus MRS ($^{31}$P-MRS) and found a decreased alpha nucleotide triphosphate ($\alpha$-NTP)/total phosphate (TP) ($p < 0.001$) ratio in patients with MASH [20] |
| SPIO | SPIO nanoparticles are a type of contrast agent used in MRI. It enhances MRI by shortening the T1 and T2/T2* relaxation times | Hepatic Kupffer cells effectively remove SPIO from the circulating blood via phagocytosis; thus, they can be imaged using hepatic signal intensity changes. As Kupffer cells are felt to be involved in the pathogenesis of MASH, the changes could be diagnostic of MASH [73] |
| *Ultrasound* | | |
| Ultrasound scores | A transducer transmits sound waves through the body. The sound waves are reflected to the transducer as echoes. The scanner then calculates the distance from the transducer to the tissue boundary and generates images of tissues based on the distance | There are several different specific parameters seen on ultrasound which have been combined in the studies to create scores that diagnose MASH. These either focus on parameters due to increased steatosis, for example, echogenicity of the liver and hepatorenal angle, which have been found to increase the likelihood of MASH. Or parameters affected by inflammation or steatosis, such as blurring of specific anatomy such as vessels, diaphragm and gallbladder [21, 25, 57]. Ultrasound can also utilise 2D SWE and other techniques as below |
| RF | RF data alters are related to microstructure changes within tissues [37] | • Mechanisms are unclear; however, Gao et al. hypothesise that the selected parameters used in their study reflect the intensity and scattering of liver ultrasound signals and are related to the microstructural changes due to liver inflammation [37] |
| Microbubbles | 2D ultrasound with microbubble contrast agent, which is phagocytosed by liver Kupffer cells, enhancing the liver parenchyma | Contrast enhancement is decreased in patients with MASH, compared to NAFLD patients. This is felt to be explained by the changes in the sinusoidal endothelial system [41] |
| ARFI | A brief acoustic radiation force impulse is passed through tissue to create displacement. It detects liver stiffness | Detects increased stiffness of the liver secondary to developing fibrosis and inflammation |
| TE | Mechanical vibration is applied over the liver to create shearwaves. The wavelength is reflective of liver stiffness | Detects increased stiffness of the liver secondary to developing fibrosis and inflammation |
| SWE | It can be 2D ultrasound shear wave or point shear wave elastography. A transducer induces an elastic shear wave in the underlying tissue. The ultrasound measures the velocity of the shear wave through the tissue. In pSWE, the measurement of stiffness is conducted by targeting a single point in a localised region of interest. This is different from 2D-SWE, in which the device calculates the mean stiffness within a larger area | SWE can detect changes in stiffness of liver tissue, which reflect increases in inflammation and fibrosis. This stiffness can be measured via three different methods<br>• Liver stiffness in kilopascals (kPa)<br>• Dispersion slope in metres/second/kilohertz ((m/s)/kHz): parameter used to measure tissue viscosity by measuring the change in shear wave speed with frequency<br>• Speed in metre per second (m/s): higher shear wave speeds are associated with higher liver stiffness |

(Continues)

**TABLE 1** | (Continued)

| | Underlying principles | Theory in diagnosing MASH |
|---|---|---|
| CAP | Measures the attenuation of an ultrasound beam as it passes through the liver | Fat affects ultrasound propagation; thus, attenuation is correlated to the percentage of steatosis. Increased steatosis could be a marker of increased risk of MASH |
| *Other* | | |
| CT | A narrow beam of X-rays is aimed at a patient and rotated around the body; this produces signals that are processed by a computer to generate cross-sectional images. The number of X-rays that pass through the object is inversely proportional to the density of the object | Several different methods<br>• Dichtel et al. and Naganawa et al. describe using different CT texture analysis methods for diagnosis [29, 61]. Alluding to a change of the texture of the liver in MASH<br>• Kim et al. describe using the liver-spleen Hounsfield unit (HU), which is a tool to look at hepatic fat content in comparison to spleen. They found it also correlated with the degree of inflammation [46] |
| Scintigraphy | A radioisotope ($^{99m}$Tc-phytate), which is a radioactive compound that emits gamma rays, is administered; this is then detected by a gamma camera. This creates images showing the distribution of the radioisotope within the body | The reduction of this radioisotope uptake during the scintigraphic exam could be a marker of MASH as $^{99m}$Tc-phytate accumulates in the liver, followed by phagocytosis by Kupffer cells. The reticuloendothelial dysfunction of Kupffer cells is involved in the pathogenesis of steatohepatitis [67] |

Abbreviations: ARFI, acoustic radiation force impulse; CAP, controlled attenuation parameter; CT, computerised tomography; GA MRI, Gadoxetic acid–enhanced MR; IVIM MRI, intravoxel incoherent motion; MpMRI, multiparametric magnetic resonance; MRE, magnetic resonance elastography; MRI-PDFF, MRI proton density fat fraction; MRS, magnetic resonance spectroscopy MRI; SPIO, Superparamagnetic iron oxide; SWE, shear wave elastography; TE, transient elastography.

patients with liver biopsies, ten had a control group that did not undergo a liver biopsy [20, 27, 31, 33, 41, 44, 48, 68, 72, 85]. However, only three studies included the non-biopsied control group in their diagnostic accuracy results [20, 41, 44]. For the paper by Lirussi et al., it was unclear if they included the control group in their results [85]. One study, Zhou et al. included 23 biopsied controls [81]. Only 14/69 studies had preset thresholds for their imaging test or undertook a validation [34, 36, 37, 42, 53, 59–61, 69, 70, 75, 81, 86, 88]. Macias et al. validated their study but only gave negative predictive value (NPV) and positive predictive value (PPV) from the study [59]. The remainder (55/69 studies) were regarded as high risk of bias for the index test category within the tool. In all, 36 studies were scored as either unclear or high risk of bias after excluding patients from analysis due to missing data, or unreliable/unachievable scan results or inadequate liver biopsies [20–24, 26, 29, 31–35, 37, 38, 41–44, 47, 49–52, 54, 58–60, 62–65, 71, 77, 83, 86, 88].

## 3.5 | Certainty of Evidence

The level of certainty of evidence, according to GRADE criteria, is reported for the different imaging methods in the Supporting Information.

## 3.6 | Data Synthesis

We grouped studies into similar imaging methods to compare. Where there was a testing and validation cohort, we reported the outcomes of the validation cohort. Results are presented in tables to the nearest two decimal points. We undertook a narrative synthesis of the data. Due to the heterogeneity of the data, different variations on imaging tests, different MASH definitions and different thresholds for tests, data pooling was not possible.

## 3.7 | MASH Diagnosis

### 3.7.1 | Magnetic Resonance Imaging (MRI)

There were 31 studies reviewing MRI and steatohepatitis presented in Table 2.

**3.7.1.1 | Non-MRE Studies.** The best AUROC within the MRI studies was 1.0 reported by Kim et al. using magnetic resonance spectroscopy (MRS). The study looked at metabolic changes using $^1$H-MRS with long echo time [48]. A larger study on MRS was by Abrigo et al., which showed a lower 0.71 AUROC [20]. Both these studies suffered from bias due to not having preset thresholds. However, the Abrigo study was at overall higher risk of bias due to including a non-biopsied control group in its results as well as missing data due to unsuccessful MRS in 8 patients [20]. Other techniques include gadoxetic acid-enhanced MRI (GA-MRI) used by Bastati et al. with an AUROC of 0.85 [26] but this was not replicated in a further study by Amorim et al. (AUROC 0.68) [24]. Apart from having no preset thresholds, both studies were at lower risk of bias. Both had to exclude a low number of patients due to motion artefacts on MRI (3 in Amorim and 10 in Bastati), though Bastati did additionally exclude 7 patients with incomplete biopsy results [24, 26]. MRI-PDFF scored AUROCs of 0.52–0.84. The highest AUROC was by Garteiser et al.; MRI-PDFF had 97% sensitivity and 60% specificity, 50% PPV and 98% NPV for diagnosing NASH at a

**TABLE 2** | Diagnostic accuracy of MRI studies in MASH diagnosis.

| Method/model used | Authors and year | AUROC (95% CI) | Population number (MASH) | Pathology definition | Sensitivity and specificity |
|---|---|---|---|---|---|
| *MRE or MRE with other techniques* | | | | | |
| 2D MRE | 1. Allen et al. 2020 [22]<br>2. Alsaqal et al. 2021 [23]<br>3. Chen et al. 2011 [28]<br>4. Costa-silva et al. 2018 [27]<br>5. Lee et al. 2020 [54]<br>6. Park et al. 2017 [63]<br>7. Troelstra et al. 2021 [75, 76]<br>8. Loomba et al. 2016 [87]<br>9. Imajo et al. 2021 [43]<br>10. Li et al. 2023 [56]<br>11. Loomba et al. 2014 [88] | 1. 0.61 (CI 0.53–0.69) cross-validated C-statistic for NASH diagnosis and NAS prediction 0.64 (CI 0.58–0.70). Both combined cohort<br>2. 0.74 (CL N/A)<br>3. 0.93 (CI N/A)<br>4. 0.79 (0.66–0.95)<br>5. 0.77 (CI 0.69–0.84)<br>6. 0.70 (CI 0.57–0.82)<br>7. 0.79 (CI N/A) MRE G1 + MRE G2 0.69 (CI N/A)<br>8. 0.75 (CI N/A)<br>9. 0.57 (CI 0.47–0.67)<br>10. 0.73 (CI N/A)<br>11. 0.73 (CI N/A) | 1. 175 (81)<br>2. 68 (53)<br>3. 58 (36)<br>4. 49 (25)<br>5. 130 (70)<br>6. 104 (72)<br>7. 37 (22)<br>8. 100 (72)<br>9. 145 (87)<br>10. 104 (45)<br>11. 117 (106) | 1. The presence of steatosis with (1) ballooning or (2) lobular inflammation with associated fibrosis<br>2. SAF grading<br>3. Brunt criteria<br>4. NAS ≥ 5<br>5. > 5% steatosis and inflammation with hepatocyte ballooning<br>6. NAS ≥ 3<br>7. NAS ≥ 3. Steatosis and ballooning grade and lobular inflammation grade ≥ 1 for each<br>8. NAS ≥ 5<br>9. NASH was defined as NAS ≥ 4 and at least ballooning ≥ 1 and inflammation ≥ 1<br>10. NASH CRN; no specific NAS<br>11. NAS ≥ 3 | 1. N/A<br>2. 0.49 se and 1.00 spec cut-off > 2.74 kPa<br>3. 0.94 se and 0.73 spec cut-off > 2.74 kPa and 0.83 se and 0.82 spec cut-off > 2.90 kPa<br>4. 0.72 (0.48–0.86) se and 0.88 (0.66–0.97) at cut-off > 3.24 kPa<br>5. N/A<br>6. 0.64 se and 0.68 spec for cut-off 2.53 kPa<br>7. 0.70 se and 0.87 spec for cut-off 2.27 kPa MRE G1 + 0.45 se and 1.00 spec for 0.88 kPa cut-off MRE G2<br>8. N/A<br>9. N/A<br>10. N/A<br>11. 0.43 se and 0.89 spec (cross-validated) at 3.26 kPa |
| 3D MRE | Loomba et al. 2016 [87] | 1. 0.76 (60 Hz) (CI N/A)<br>2. 0.74 (40 Hz) (CI N/A) | 100 (72) | NAS ≥ 5 | 0.67 se and 0.80 spec for 3D MRE model |
| 3D-MRE + MRI-PDFF | Allen et al. 2020 [22] | 0.73 (CI 0.65–0.81) cross-validated C-statistic for NASH diagnosis and 0.82 (CI 0.77–0.87) for predicting the NAS score | 175 (81) | The presence of steatosis with (1) hepatocellular ballooning or (2) lobular inflammation with associated fibrosis | 0.67 se and 0.80 spec for 3D MRE model |
| 2D MRE + MRI-PDFF | 1. Alsaqal et al. 2021 [23]<br>2. Li et al. 2023 [56] | 1. 0.84 (CI N/A)<br>2. 0.79 (CI N/A) | 1. 68 (53)<br>2. 104 (45) | 1. SAF grading<br>2. NASH CRN. No specific number | 1. 0.74 se and 0.87 spec cut-off > 1.38 kPa<br>2. N/A |

(Continues)

**TABLE 2** | (Continued)

| Method/model used | Authors and year | AUROC (95% CI) | Population number (MASH) | Pathology definition | Sensitivity and specificity |
|---|---|---|---|---|---|
| Machine learning model with MRE + MRI-PDFF | Dzyubak et al. 2021 [30] | 0.87 (CI N/A) for NASH diagnosis + 0.85 (CI N/A) cross-validated C-statistic of for both manual and automated measurements for model to predict NAS score | 83 (34) | NAS score, does not specify a number | N/A |
| MRE-based NASH score with demographic and laboratory variables | Lee et al. 2022 [53] | 0.84 (CI 0.77–0.91) + C-statistic 0.81 at bootstrapping | 127 (69) | Steatosis was ≥ 5% with inflammation and ballooning of hepatocytes | 0.68 se and 0.91 spec for cut-off 0.68 0.91 se and 0.55 spec for cut-off 0.37 |
| MRE + T1 mapping | Li et al. 2023 [56] | 0.74 (CI N/A) | 104 (45) | NASH CRN. No specific number | N/A |
| MRI-PDFF + T1 + MRE | Li et al. 2023 [56] | 0.83 (CI 0.75–0.92) | 104 (45) | NASH CRN. No specific number | N/A |
| *Non-MRE-based MRI scores* | | | | | |
| MRI-PDFF | 1. Alsaqal et al. 2021 [23] 2. Troelstra et al. 2021 [76] 3. Imajo et al. 2021 [43] 4. Garteiser et al. 2021 [38] 5. Idilman et al. 2021 [40] 6. Li et al. 2023 [56] 7. Wildman-Tobriner et al. 2018 [77] | 1. 0.70 (CI N/A) NASH 2. 0.57 (CI N/A) PDFF MRI-M + PDFF MRS 0.56 (CI N/A) + PDFF Dixon 0.52 (CI N/A) 3. 0.80 (CI: 0.73–0.87) 4. 0.84 (CI 0.77–0.91) 5. 0.64 (CI 0.54–0.73) 6. 0.82 PDFF Dixon (CI 0.72–0.91) 7. 0.72 (CI 0.59–0.84) | 1. 68 (53) 2. 37 (22) 3. 145 (87) 4. 152 (30) 5. 135 (73) 6. 104 (45) 7. 170 (51) | 1. SAF 2. Steatosis and ballooning grade was ≥ 1 and lobular inflammation grade ≥ 1 was present 3. NAS ≥ 4 and at least one in ballooning ≥ 1 and inflammation ≥ 1 4. FLIP criteria 5. NASH CRN. No specific number 6. NASH CRN. No specific number 7. NAS ≥ 4. Combined trial cohorts of 3 trails but uses only two trails for NASH diagnosis. One trial NAS ≥ 4 and ≥ 1 hepatocyte ballooning and other trial NAS ≥ 4 with ≥ 1 hepatocyte ballooning, inflammation and steatosis. Fibrosis stage 1–3 was needed for enrolment | 1. 0.81 se and 0.60 spec 2. 0.41 se and 0.87 spec for 23.72% cut-off + 0.46 se and 0.73 spec for 21.19% cut-off + 0.33 se and 0.87 spec for 21.84% cut-off 3. N/A 4. 0.97 (CI 0.86–1.00) se and 0.60 (CI 0.50–0.71) spec at 12% cut-off 5. 0 0.61 (0.49–0.72) se and 0.65 (0.51–0.76) spec 6. N/A 7. 0.75 (CI 0.35–0.97) se and 0.75 (CI 0.67–0.81) spec cut-off 12.4% |

(Continues)

TABLE 2 | (Continued)

| Method/model used | Authors and year | AUROC (95% CI) | Population number (MASH) | Pathology definition | Sensitivity and specificity |
|---|---|---|---|---|---|
| T1 mapping and MRI PDFF | 1. Imajo et al. 2021 [43]<br>2. Li et al. 2023 [56] | 1. 0.83 (CI: 0.76–0.90)<br>2. 0.81 (CI N/A) | 1. 145 (87)<br>2. 104 (45) | 1. NAS ≥4 and at least one in ballooning ≥1 and inflammation ≥1<br>2. NASH CRN. No specific number | 1. N/A<br>2. N/A |
| T1 mapping | 1. Eddowes et al. 2018 [31]<br>2. Li et al. 2023 [56] | 1. 0.69 (0.50–0.88) Simple steatosis and NASH 0.74 (0.59–0.88) NAS <5 versus NAS ≥5<br>2. 0.72 (CI N/A) | 1. 50 (38)<br>2. 104 (45) | 1. Presence of lobular inflammation and hepatocyte ballooning. Secondly, NASH CRN score was used for grading<br>2. NASH CRN. No specific number | 1. N/A |
| M-PAST – MRI-PDFF and AST | Imajo et al. 2023 [42] | 0.81 (CI 0.73–0.86) validation cohort In the UCSD cohort (non-Japanese cohort), M-PAST (0.77) (CI 0.71–0.83) | 169 (98) validation cohort 234 (117) USCD cohort | NAS ≥4 | N/A |
| IVIM | 1. Troelstra et al. 2021 [76]<br>2. Parente et al. 2015 [83] | 1. 0.73 (CI N/A) IVIM D+IVIM D* 0.58 (CI N/A)+IVIM f 0.68 (CI N/A)<br>2. 0.74 (CI N/A) D+D* 0.69 (CI N/A)+F 0.61 (CI N/A) | 1. 37 (22)<br>2. 59 (22) | 1. Steatosis and ballooning grade was ≥1 and lobular inflammation grade≥1 was present<br>2. Steatosis and mixed-cell inflammatory infiltrates in the hepatic lobules and damage of hepatocytes. Damage characterised by contiguous patch of hepatocytes showing prominent ballooning with or without apoptosis and necrosis | 1. 0.86 se and 0.53 spec for 0.0012 mm²/s cut-off + 0.81 se and 0.53 spec for 22.90% cut-off + 0.48 se and 0.80 spec for 0.10 mm2/s cut-off<br>2. D 0.69 se and 0.66 spec cut-off 0.76<br>3. D* 0.69 se and 0.71 spec cut-off 0.42<br>4. F 0.49 se and 0.70 spec cut-off 0.34 |
| NASH MRI – model based on independent MRI variables | Gallego-Duran et al. 2016 [36] | 0.83 (CI 0.73–0.93) Validation cohort | 87 (44) Validation cohort | Patterns of histological distribution focusing on inflammation and ballooning | 0.87 se and spec 0.60 cut-off 0.5 |
| Magnetic susceptibility | Leporq et al. 2017 [55] | 0.91 (CI 0.79–1.0) | 32 (20) Validation cohort | FLIP criteria | N/A |

(Continues)

**TABLE 2** | (Continued)

| Method/model used | Authors and year | AUROC (95% CI) | Population number (MASH) | Pathology definition | Sensitivity and specificity |
|---|---|---|---|---|---|
| MR – MASH score – MRI PDFF and waist circumference (in MR images) and height | Marti-Aguado et al. 2023 [60] | 0.86 (CI 0.77–0.92) Validation cohort | 63 (59) validation cohort | NAS ≥4 presence of steatosis, lobular inflammation and ballooning with or without fibrosis | 0.91 se and 0.73 spec (rule out MASH) (MR–MASH score <400) 0.56 se and 0.87 spec (rule in MASH) (MR–MASH score ≥600) |
| LiverMultiScan measuring LIF scores from $T_1$ and $T_2$ | Pavlides et al. 2017 [64] | 0.80 (CI 0.69–0.92) | 71 (46) | FLIP criteria | 0.91 (0.79–0.98) se and 0.52 (0.31–0.72) spec 1.4 LIF cut-off |
| MRI parameters | Saadeh et al. 2002 [70] | None of the radiological features included could distinguish patients with the pathologic diagnosis of NASH from steatosis | 25 (17) | Type 3 (steatosis, lobular inflammation, hepatocyte ballooning) or type 4 (steatosis, lobular inflammation, hepatocyte ballooning and Mallory's hyaline or fibrosis) | N/A |
| Multiparametric MR index | Kim et al. 2020 [82] | 0.88 (CI N/A) | 47 (20) | SAF criteria | 0.95 se and 0.78 spec cut-off of 4.6 + 0.80 se and 0.85 spec cut-off of 6 |
| GA-MRI | 1. Amorim et al. 2020 [24] 2. Bastati et al. 2014 [26] | 1. 0.68 (CI N/A) 2. 0.85 (0.75–0.91 CI) | 1. 56 (38) 2. 81 (35) | 1. NASH CRN, does not specify a NAS score 2. SAF – Activity score >2 and any fibrosis/steatosis | 1. 0.47 se and 0.83 CEI cut-off 1.76 + 0.92 se and 0.28 spec CEI cut-off 2.00 + 0.32 se and 0.94 spec CEI cut-off 1.66 2. 0.97 (CI 84.7–99.9) se and 0.63 (CI 47.5–76.8) spec cut-off 1.24 |
| MRS | 1. Abrigo et al. 2014 [20] 2. Kim et al. 2017 [48] | 1. 0.71 (0.62–0.79 CI) 2. 1.00 (CI 1.00–1.00) for Ala + 0.78 (CI 0.61–0.96) for Lac + TG | 1. 151 (91) 2. 32 (11) | 1. Matteoni classification – type 3 and type 4 were defined as NASH 2. NAS ≥ 5 | 1. 0.91 se and 0.16 spec α-NTP/TP cut-off of ≤16.36% +0.28 se and 0.91 spec cut-off of ≤10.57% 2. 1.00 se and 1.00 spec cut-off value >16.04% Ala 0.82 se and 0.67 spec at a cut-off value >360.8% Lac + TG |

(Continues)

**TABLE 2** | (Continued)

| Method/model used | Authors and year | AUROC (95% CI) | Population number (MASH) | Pathology definition | Sensitivity and specificity |
|---|---|---|---|---|---|
| SPIO-enhanced | 1. Tomita et al. 2008 [73]<br>2. Smits et al. 2016 [68] | 1. 0.79 (CI 0.56–1.01) T values + %T2 values 0.83 (CI 0.64–1.02)<br>2. 0.87 (CI 0.72–1.00) | 1. 19 (10)<br>2. 24 (13) | 1. NAS ≥ 5<br>2. NAS ≥ 5 when steatosis, inflammation and ballooning were present | 1. Se 1.0 and 0.67 spec cut-off tau value of 42.8 + 0.88 se and 0.73 spec cut-off %T2 value of 32.5<br>2. 0.77 se and 0.91 spec at ΔR2* <45.5s⁻¹ cut-off 0.92 se and 0.55 spec cut-off <59.2.s⁻¹ |

Abbreviations: AST, aspartate aminotransferase; CEI, contrast enhancement index; CI, Confidence Interval; FLIP, fatty liver inhibition of progression; GA-MRI, gadoxetic acid-enhanced MRI; IVIM D*, pseudodiffusion coefficient; IVIM D, diffusion coefficient; IVIM, intravoxel incoherent motion; kPa, kilopascals; LIF, liver inflammation and fibrosis score; mm²/s, millimetres per second; MRE, magnetic resonance elastography; MRE G1, shear modulus; MRE G2, loss modulus; MRS, magnetic resonance spectroscopy; NAS, NAFLD activity score; PDFF Dixon, three-point DIXON proton density fat fraction; PDFF MRS, magnetic resonance spectroscopy proton density fat fraction; PDFF, proton density fat fraction; PDFF-M, magnitude-based MRI proton density fat fraction; SAF, steatosis activity and fibrosis score; Se, sensitivity; Spec, specificity; SPIO, Superparamagnetic iron oxidesg.

cut-off of 12% [38]. These results suggest that MRI-PDFF is better at MASH exclusion rather than diagnosis. This high sensitivity was not reflected in the other studies with lower sensitivities, even in the study by Wildman-Tobriner et al., who used a similar cut-off at 12.4%, reporting a sensitivity of 75% [77]. Troelstra et al. conducted PDFF with three different MRI techniques, such as MRS, Dixon and magnitude-based, but achieved the lowest AUROCs among the studies [76]. MRI-PDFF scored higher AUROCs when combined with other modality such as MRE in the study by Dzyubak et al. reporting an AUROC of 0.87 for NASH diagnosis and a cross-validated C-statistic of 0.85 for NAS prediction with additional use of machine model learning analysis of the imaging data. However, this was in a group of bariatric surgery patients, making this study at risk of selection bias [30]. A few other studies combining 2D MRE and MRI-PDFF scored lower AUROCs of 0.84 [23] and 0.79 [56], and the 3D MRE and MRI-PDFF combination resulting in 0.73 [22]. This latter study by Allen *et al.* in a bariatric cohort again shows better results for NASH exclusion than diagnosis [22]. Leporq et al. reported on a decrease in magnetic susceptibility in MASH using MRI; the study achieved an AUROC of 0.91. Apart from no preset thresholds, the paper otherwise scored low risk of bias; however, it was unclear from the paper whether the pathologist was blinded to the results of the MRI [55].

**3.7.1.2** | **MRE Studies.** The second best AUROC overall for MRI studies was reported by Chen et al. at 0.93 in MRE. This was using Brunt criteria as the definition of NASH, discriminating patients with NASH from those with simple steatosis in a population of 58 NAFLD patients [28]. Other studies reporting on MRE alone report AUROC ranging from 0.57 to 0.87. The lowest score was a study by Imajo et al. Its pathology definition did not include fibrosis [43]. If we use the cut-off for definite NASH via the NASH CRN definition (NAS ≥ 5), then the highest AUROC by study is by Costa-silva et al., who report an AUROC of 0.79 for MRE [27]. The studies by Chen and Costa-Silva, apart from no preset thresholds, scored otherwise low for bias [27, 28]. The study by Imajo, however, scored high risk of bias due to no preset thresholds, and it was unclear if the pathologist was blinded (although they used 3 different pathologists to score the biopsies). There was a large volume of missing data in this paper, but it only had one missing MRE test [43].

### 3.7.2 | Ultrasound

**3.7.2.1** | **Ultrasound, Shear Wave (SW), Acoustic Radiation Force Impulse (ARFI) and MASH.** There are 27 ultrasound-based studies: five did not report AUROC, including one because it didn't reach statistical significance to build a model (Table 3). Within the studies, the AUROC ranged from 0.65–0.94. The highest AUROC for ultrasound studies was 0.94 jointly by two studies. Firstly, by Funada et al., they developed a model called the NASH pentagon, which included five parameters: SW speed (SWS), SW dispersion slope (SWD), attenuation imaging value (also named attenuation coefficient (AC)), Fib-4 and BMI [35]. Secondly by Jang et al., who created a risk score based on ultrasound, AC and DS [44]. Funada et al.'s test suffered bias as they had no preset thresholds and not everyone in the study got a biopsy, although they only used biopsied patients in the results

[35]. Jang also suffered from no preset thresholds, and it was unclear if the pathologist was blinded. They also included living donors. However, in the Supporting Information, excluding patients for living donation, the AUROC was 0.93 (95% CI: 0.84–0.98) for the risk score [44]. All the studies with SW elastography (SWE) utilised 2D-SWE rather than point-SWE (pSWE). The studies with the highest sensitivity (100%) and specificity (100%) were by Lijima et al. They employed the use of a microbubble contrast agent; although the study did show a decrease in uptake of contrast agents for patients with NASH, there was no correlation to the degree of fibrosis or steatosis. The study did not report an AUROC, and we found no other similar study [41].

Three studies used a model called the ultrasound fatty liver indicator (US-FLI) developed by Ballestri et al., based on five ultrasound parameters – liver/kidney contrast, posterior attenuation of ultrasound beam, vessel blurring, gallbladder wall blurring, diaphragm blurring and areas of focal sparing – to give a score of 2–8 [25, 69, 84]. Nelson et al. reported an AUROC of 0.93 (CI 0.83–0.97) when the US-FLI $\geq 5$ to differentiate normal versus NASH [84]. Ballestri et al. demonstrated an AUROC of 0.80 (CI 0.676–0.916) in their study on patients with a NAS $\geq 5$. It scored high risk for patient selection as it excluded insulin-treated type 2 diabetes patients [25]. A further study using point of care ultrasound (POCUS) also used US-FLI; they demonstrated an AUROC of 0.78 [69].

Seven studies in total evaluated SW, reporting low to high AUROCs: 0.58 to 0.95 [43, 44, 49, 50, 62, 71, 81]. The highest AUROC was reported in a study by Zhou et al., who used the Brunt classification of NASH and therefore included fibrosis in this, and the AUROC of 0.95 was seen in a subset of the population with cirrhosis. It suffered from bias due to including biopsied controls in its cohort; however, it did use 5-fold cross-validation for its results [81].

We found two studies reporting on ARFI alone; these report good AUROC of 0.87 and 0.90 and use similar cut-offs (ARFI velocity $> 1.10$ m/s and 1.3 m/s). They both report on high sensitivities and specialities [33, 39].

**3.7.2.2 | Vibration-Controlled Transient Elastography (VCTE), CAP and MASH.** Within the 11 studies that evaluated TE, AUROC ranged from 0.35 to 0.82 (Table 4). The lowest scoring was for Park et al. [63]. It included patients with borderline NASH or definite NASH as per the NASH CRN within its pathological definition. The highest AUROC was shared by two studies: firstly by Salvati et al. for a model combining TE and BMI score [66], and secondly by Eddowes et al., which categorised NASH based on the presence of lobular inflammation and hepatocyte ballooning. Within the same study, when they also report on NAS $\geq 5$, the AUROC was 0.74 (CI 0.59–0.89) [31]. There was a wide range of thresholds used for TE from $> 5.3$ to $> 9.9$ kpa used, and five studies did not report on the threshold used.

One study by Imajo et al. compared SW on 2D ultrasound and Fibroscan. Fibroscan was reported to have a 0.56 AUROC for NASH, and 2DSW was reported to have an AUROC of 0.58 (CI 0.45–0.71). In the same study, the authors also combined Fibroscan with CAP for NASH diagnosis, which improved the AUROC to 0.71 (CI: 0.66–0.8) [43]. There was no other paper using this combination. Kim et al. also combined TE with US features of 2D-SW and HRI for diagnosis of NAS $\geq 5$. This gave an AUROC of 0.79 (CI 0.66–0.88); AST was added to this model, which relived a similar AUROC of 0.80 (CI 0.67–0.89) [47].

Eight studies looked at CAP and MASH diagnosis with a range of AUROC from 0.65 to 0.93 (Table 4). The best was for the CIR score, which combined CAP score, HOMA and insulin resistance by Macias et al. in their testing cohort but does not give AUROC for their validation cohort [59]. The lowest AUROC (0.65) was by Troelstra et al. for differentiating steatosis versus NASH using CAP alone. They used a higher cut-off of $> 336$ m/s than used in other studies [76]. A cut-off $> 301$ m/s by Trowell et al. give the best AUROC of 0.80 for CAP alone [75].

### 3.7.3 | Other Imaging and MASH

There were five CT studies; these showed a range of AUROC from $< 0.60$ to 0.94 (Table 5). In the 0.94 study, this included patients with borderline NASH, and this was in patients without suspicion of fibrosis determined by low hyaluronic acid levels. When looking at patients with suspicion of fibrosis, the AUROC was 0.60 in the same study [61]. The second highest AUROC was 0.90 when looking at the mean value of positive pixels (MPP) below 54.0 in a study by Dichtel et al. [29] This also reported good sensitivity (100%) and specificity (100%).

There were scintigraphy studies [45, 67]. In the study by Kikuchi et al., scintigraphy images were obtained after the infusion of radiopharmaceutical $^{99m}$Tc-phytate in liver/spleen, spleen/heart, liver/heart ratios and $^{99m}$Tc-isonitrile liver/heart ratio. The authors saw a reduction in spleen/uptake ratios with NASH versus steatosis [45]. This was not replicated in a later study by D'Avila da Silva et al., which did not show any difference in liver/spleen uptake between groups [67].

## 4 | Clinical Trials

We identified eight clinical trials in our search (Supporting Information) [82, 91–97]. One is already finished and included in our paper [82]. They evaluate 2D SW, MRI liver multiscan, PET, ultrasound and spin-lock MRI. Four trials have finished recruiting and two have yet to start recruiting. One large trial, with a target recruitment of 450 patients, called the LITMUS imaging study is evaluating Liver*MultiScan*, MRE, MRI, MRI PDFF, 2D SWE, pSWE and TE. Its primary outcome is the diagnosis of fibrosis, but has a MASH diagnosis in its secondary outcome [96].

## 5 | Discussion

This study critically reviewed 69 papers reporting an imaging test or an imaging-based score for the diagnosis of MASH. We searched a range of databases to include grey literature and trial registers. We also report on the use of pathology definitions to make it transparent what was being tested and included all imaging tools found rather than focusing on one

**TABLE 3** | Diagnostic accuracy of ultrasound, SW and ARFI in MASH diagnosis.

| Model used | Authors and year | AUROC (95% CL) | Population number (MASH) | Path grading | Sensitivity and specificity |
|---|---|---|---|---|---|
| *Model using 2D US* | | | | | |
| US-FLI based on liver/kidney contrast, posterior attenuation of ultrasound beam, vessel blurring, gallbladder wall blurring, diaphragm blurring and areas of focal sparing | 1. Ballestri et al. 2012 [25]<br>2. Nelson et al. 2020 [84]<br>3. Sourianarayanane et al. 2023 (POCUS) [69] | 1. 0.80 (CI 0.676–0.916)<br>2. 0.93 (CI 0.83–0.97) when the US-FLI ≥5 in normal versus NASH + for steatosis versus NASH was 0.65 (CI 0.56–0.72) when the US-FLI ≥5<br>3. 0.78 (CI N/A) + When weighted for the effects of BMI, 0.71 (CI N/A) | 1. 53 (35)<br>2. 208 (105)<br>3. 113 (78) 82 had POCUS scans | 1. NAS > 5 was considered compatible with the diagnosis of severe NASH But ≥ 5 used for AUROC<br>2. Brunt classification<br>3. NASH CRN. No specific NAS number given | 1. Se N/A and 0.46 spec at US FLI <4<br>2. 0.7 se and 0.93 spec US-FLI score≤4+ US-FLI score≥5 0.93 se and 0.57 spec in normal versus NASH + in steatosis versus NASH US-FLI ≤4 Se 0.91 and spec 0.18+ US-FLI score≥5 Se 0.78 and spec 0.56<br>3. 0.81 se and 0.63 spec at cut-off USFLI ≥6+USFLI ≤3, 1.0 ruling out NASH. When weighted for the effects of BMI 0.82 se and spec 0.51 |
| Attenuation coefficient | 1. Jang et al. 2022 [44]<br>2. Lee et al. 2021 [50]<br>3. Sugimoto et al. 2020 [71] | 1. 0.84 (CI 0.76–0.90)<br>2. 0.84 (CI 0.75–0.90)<br>3. 0.71 (CI 0.58–0.83) | 1. 132 (53)<br>2. 102 (55)<br>3. 111 (90) | 1. NAS >3 + ≥1 point for each category – steatosis, inflammation, ballooning<br>2. NAS >3 + ≥1 point for each category – steatosis, inflammation, ballooning<br>3. Matteoni classification – type 3 and type 4 were defined as NASH | 1. 0.81 se and 0.80 spec<br>2. 0.78 se and 0.85 spec<br>3. N/A |
| Echogenicity of liver, visualisation of blood vessel, diaphragm and posterior right lobe liver | Manasra et al. 2021 [21] | N/A | 45 (19) | ≥5 NAS with presence of steatosis, inflammation and ballooning. Or NAS 3–4 with grade 1 steatosis and ballooning | 1.0 se and 0.28 spec |
| Antwerp NASH severity score (ALT, US echogenicity and posterior beam attenuation and C-peptide) | Francque et al. 2012 [34] | 0.88 (CI N/A) Validation cohort | 313 (61) | NAS ≥5 | 0.96 se and 0.46 spec |

(Continues)

**TABLE 3** | (Continued)

| Model used | Authors and year | AUROC (95% CL) | Population number (MASH) | Path grading | Sensitivity and specificity |
|---|---|---|---|---|---|
| US + 'NASH pentagon' (five parameters of 2D SW (speed), SW (DS), attenuation imaging value, Fib-4 index and BMI | Funada et al. 2023 [35] | 0.94 (CI N/A) Biopsied cohort | 31 (25) biopsied cohort | Comprehensive diagnosis to include macrovesicular steatosis, ballooning degeneration of hepatocytes, inflammation and apoptotic bodies in the lobules, Mallory-Denk bodies and fibrosis | N/A |
| US + Radiofrequency | Gao et al. 2022 [37] | 0.82 (CI 0.73–0.91) Validation cohort | 259 (100) total cohort and 87 validation cohort | NAS ≥ 5 | 0.87 se and 0.67 spec at cut-off QUS ≤ 6.0 0.38 se and 0.90 spec at cut-off QUS ≥ 7.5 |
| US+ microbubble contrast | Lijima et al. 2007 [41] | N/A | 76 (21) | Brunt criteria | 1.0 se and 1.0 spec at cut-off 43.6 signal intensity at 20 min 1.0 se and 0.95 spec at cut-off 137.8 Signal intensity at 5 min |
| US PATT | Lirussi et al. 2009 [85] | 0.75 (CI N/A) | 108 (27) | NAS ≥ 3 borderline NASH | 1.0 se and 0.5 spec cut-off value of PATT of 11.8 mm |
| US + AC and DS risk score | 1. Jang et al. 2022 [44] 2. Torkzaban et al. 2023 [74] | 1. 0.94 (CI 0.89–0.98) in total cohort and 0.93 (95% CI: 0.84–0.98). in cohort excluding living donors 2. 0.69 (CI N/A) | 1. 132 (53) 2. 34 (11) | 1. NAS ≥ 3 + ≥ 1 point for each category – steatosis, inflammation, ballooning 2. NAS ≥ 3. One point for each category | 1. 0.81 se and 0.96 spec in total cohort and 0.85 se and 0.96 spec in cohort excluding living donors 2. N/A |
| 1. Grey-scale US index (SWE and HRI) and AST 2. Grey-scale US index (SWE + HRI) 3. Grey-scale US index (SWE + HRI) + fibroscan 4. Grey-scale US index (SWE + HRI) + fibroscan + AST | Kim et al. 2022 [47] | 1. 0.81 (CI 0.68–0.90) 2. 0.79 (CI 0.67–0.89) 3. 0.79 (CI 0.66–0.88) 4. 0.80 (CI 0.67–0.89) | 60 (25) | NAS ≥ 5 | 1. 68.0 (46.5–85.1) se and 80.0 (63.1–91.6) 2. 68.0 (46.5–85.1) se and 85.7 (69.7–95.2) spec 3. 72.0 (50.6–87.9) se and 85.7 (69.7–95.2) spec 4. 56.0 (34.9–75.6) se and 97.1 (85.1–99.9) spec |

(Continues)

**TABLE 3** | (Continued)

| Model used | Authors and year | AUROC (95% CL) | Population number (MASH) | Path grading | Sensitivity and specificity |
|---|---|---|---|---|---|
| US + scoring system based on AC and SWE | Lee et al. 2021 [50] | 1. 0.93 (CI 0.87–0.97) | 102 (55) | NAS > 3 + ≥1 point for each category – steatosis, inflammation, ballooning | 1. 0.93 se and 0.83 spec |
| 1. Modified Fatty Score (FS) (0–2)– 0 for FS<7 and the sum of parenchymal echogenicity + gallbladder wall blurring <3; score 1 for FS ≥7 or the latter ≥3; score 2 for FS ≥7 and the latter ≥3<br>2. Fatty score (0–8) – 5 US parameters parenchymal echogenicity, far gain attenuation, gallbladder wall blurring, portal veinwall blurring and hepatic vein blurring | Liang et al. 2007 [57] | 1. 0.82 (CI N/A)<br>2. 0.79 (CI N/A) | 101 (72) | Brunt criteria with modifications. NASH was defined as the presence of fibrosis (grade 1 or higher) or acinar zone 3 hepatocellular injury with ballooning (grade 2 or higher) | 1. N/A<br>2. 0.72 se and 0.86 spec at cut-off 3<br>3. 0.81 se and 0.66 spec at cut-off 7 |
| US-Fatty liver (mild, moderate, or severe according to the fall in echo amplitude, extent of liver/kidney discrepancy and of echo loss from portal vein) | Petrick et al. 2015 [86] | N/A | 513 (164) | Brunt criteria | Mild 0.89 se and 0.45 spec Moderate 0.61 se and 0.70 spec |
| Total Cholesterol, ALT, Ast/alt ratio, γGT and US showing steatosis (ALT =/> 30) | Pulzi et al. 2011 [65] | 0.82 (CI 0.67–0.98) | 45 (10) | Brunt criteria. Patients with steatohepatitis or fibrosis | 0.70 se and 0.89 spec |
| Three US parameters; DS, AC and SWS | 1. Sugimoto et al. 2020 [71]<br>2. Torkzaban et al. 2023 [74] | 1. 0.81 (CI 0.71–0.91)<br>2. 0.87 (CI N/A) | 1. 111 (90)<br>2. 34 (11) | 1. Matteoni classification – type 3 and type 4 were defined as NASH<br>2. NAS ≥ 3. One point for each category | 1. N/A<br>2. N/A |
| Ultrasonographic spleen longitudinal diameter | Tarantino et al. 2009 [72] | 0.92 (CI N/A) | 83 (40) | Kleiner no further specified NASH definition | 0.88 se and 0.95 spec cut-off 116 mm |
| 1. AC, LS and DS with BMI<br>2. DS + AC + BMI | Torkzaban et al. 2023 [74] | 1.0.87 (CI N/A)<br>2.0.72 (CI N/A) | 34 (11) | NAS ≥ 3. One point for each category | N/A |

(Continues)

**TABLE 3** | (Continued)

| Model used | Authors and year | AUROC (95% CL) | Population number (MASH) | Path grading | Sensitivity and specificity |
|---|---|---|---|---|---|
| 1. Attenuation of the echo amplitude, enlarged splenic diameter and presence of focal fat sparing<br>2. Attenuation of the echo amplitude and focal fat sparing | Zardi et al. 2011 [79] | N/A | 94 (74) | The appearance of steatosis, lobular and hepatocyte ballooning alone or together with Mallory's hyaline bodies or fibrosis | 1. 0.74 se and 0.66 spec at cut-off ≥5<br>2. 0.92 se and 0.75 spec at cut-off 1 |
| US + SWS, AC, NLV (normalised local variance), age and glucose | Zhao et al. 2023 [80] | 0.78 (CI N/A) Validation set | 75 (16) total cohort, validation 13 | Matteoni classification – type 3 and type 4 were defined as NASH | N/A |
| US | Saadeh et al. 2002 [70] | None of the radiological features included in the CT, US, or MRI protocols could distinguish patients with the pathologic diagnosis of NASH from steatosis | 25 (17) | Type 3 (steatosis, lobular inflammation, hepatocyte ballooning) and type 4 (steatosis, lobular inflammation, hepatocyte ballooning and Mallory's hyaline or fibrosis). Patients with 3 and 4 were categorised as NASH | N/A |
| **2D-SW** | | | | | |
| SWE (Liver stiffness) | 1. Jang et al. 2022 [44]<br>2. Zhou et al. 2022 [81]<br>3. Imajo et al. 2021 [43] | 1. 0.85 (CI 0.78–0.91)<br>2. 0.88 (CI 0.82–0.95) non-cirrhosis NASH + differentiating cirrhosis NASH 0.95 (CI 0.89–1.00) AUROC was only given for training and not testing cohort<br>3. 0.58 (CI 0.45–0.71) | 1. 132 (53)<br>2. 139 total cohort (78 patients had non-cirrhosis NASH, 17 had cirrhosis NASH)<br>3. 145 (87) | 1. NAS >3 + ≥1 point for each category – steatosis, inflammation, ballooning<br>2. Kleiner/Brunt criteria classification to include fibrosis<br>3. NAS ≥4 and at least one in ballooning ≥1 and inflammation ≥1 | 1. 0.83 se and 0.075 spec at 6 kPa<br>2. 0.71 se and 0.99 spec at cut-off 7.3 kPa Non-cirrhosis training cohort 0.86 se and 0.98 testing cohort 0.88 se and 0.91 at cut-off 12.6 kPa in the cirrhosis training cohort and 0.87 se and 0.91 spec in testing cohort<br>3. N/A |

(Continues)

**TABLE 3** | (Continued)

| Model used | Authors and year | AUROC (95% CL) | Population number (MASH) | Path grading | Sensitivity and specificity |
|---|---|---|---|---|---|
| SWS (speed) | 1. Sugimoto et al. 2020 [71] <br> 2. Ozturk et al. 2020 [62] <br> 3. Kim et al. 2022 [49] | 1. 0.70 (CI 0.58–0.83) <br> 2. 0.67 (CI 0.55–0.79) <br> 3. 0.78 (CI 0.60–0.97) | 1. 111 (90) <br> 2. 116 (24) <br> 3. 25 (13) | 1. Matteoni classification – type 3 and type 4 were defined as NASH. NAS ≥ 5 <br> 2. NAS > 3 the presence of steatosis (>5% of hepatocytes) and the presence of lobular inflammation and hepatocyte ballooning ≥1 point each | 1. N/A <br> 2. Se 0.71 and spec 0.64 cut-off 8.5 kPA (1.68 m/s) <br> 3. SW speed 0.92 se and 0.58 spec threshold 1.52 m/s |
| SWD (dispersion slope) | 1. Jang et al. 2022 [44] <br> 2. Lee et al. 2021 [50] <br> 3. Kim et al. 2022 [49] <br> 4. Sugimoto et al. 2020 [71] | 1. 0.86 (CI 0.79–0.92) <br> 2. 0.89 (CI 0.82–0.95) <br> 3. 0.83 (CI 0.66–1.00) <br> 4. 0.76 (CI 0.63–0.88) | 1. 132 (53) <br> 2. 102 (55) <br> 3. 25 (13) <br> 4. 111 (90) | 1. NAS > 3 + ≥1 point for each category – steatosis, inflammation, ballooning <br> 2. NAS > 3 + ≥1 point for each category – steatosis, inflammation, ballooning <br> 3. NAS > 3 + the presence of steatosis (> 5%) and the presence of lobular inflammation and hepatocyte ballooning ≥1 point each <br> 4. Matteoni classification – type 3 and type 4 were defined as NASH | 1. 0.85 se and 0.82 spec at 11.4 (m/s)/KHz <br> 2. 0.86 se and 0.87 spec <br> 3. se 0.69 and spec 0.92 slope 11.05 (m/s)/kHz <br> 4. N/A |
| *AFRI* | | | | | |
| ARFI | 1. Braticevici et al. 2013 [33] <br> 2. Guzmán-Aroca et al. 2012 [39] | 1. 0.87 (CI 0.78–0.95) <br> 2. 0.90 (CI N/A) Simple steatosis (group A) versus NASH or fibrosis (group B + C) | 1. 64 (43) <br> 2. 32 (group B inflammation 18, group C fibrosis 6) | 1. Brunt (Kleiner) <br> 2. Matteoni classification. group 1, steatosis without inflammation or fibrosis; group 2, steatosis with lobular inflammation without fibrosis; group 3, additional presence of ballooning hepatocytes; group 4, presence of Mallory's hyaline and/or fibrosis. The patients were classified into three groups: group A, simple steatosis; group B, inflammation; and group C, fibrosis | 1. ARFI velocity > 1.10 m/s as the cut-off 0.77 se and 0.72 spec <br> 2. Cut-off value 1.3 m/s; se 0.85 and spec 0.83 |

Abbreviations: AC, attenuation coefficient; ALT, alanine aminotransferase; ARFI, acoustic radiation force impulse; AST, aspartate aminotransferase; BMI, body mass index; CI, Confidence Interval; DS, dispersion slope; FLI, fatty liver indicator; HRI, hepatorenal index; kHz, kilohertz; kPa, kilopascals; LS, liver stiffness; m/s, metres per second; NAS, NAFLD activity score; PATT, perihepatic adipose tissue thickness; POCUS, point of care ultrasound; QUS, quantitative ultrasound; Se, sensitivity; Spec, specificity; SW, shear wave; SWE, shear wave elastography; US, ultrasound.

**TABLE 4** | Diagnostic accuracy of VCTE and CAP in MASH diagnosis.

| Method/model used | Authors and year | AUROC (95% CL) | Population number (MASH) | Path grading | Sensitivity and specificity |
|---|---|---|---|---|---|
| VCTE | 1. Alsaqal et al. 2021 [23]<br>2. Eddowes et al. 2018 [31]<br>3. Kim et al. 2022 [47]<br>4. Imajo et al. 2021 [43]<br>5. Lee et al. 2022 [52]<br>6. Lee et al. 2020 [54]<br>7. Troelstra et al. 2021 [76]<br>8. Park et al. 2017 [63]<br>9. Lee et al. 2016 [51]<br>10. Salvati et al. 2021 [66]<br>11. Eilenberg et al. 2021 [32] | 1. 0.77 (CI N/A)<br>2. 0.82 (CI 0.69–0.94) NASH versus simple steatosis + 0.74 (0.59–0.89) NAS < 5 versus NAS ≥ 5<br>3. 0.69 (CI 0.55–0.80)<br>4. 0.56 (CI 0.45–0.66)<br>5. 0.72 (C: 0.65–0.78)<br>6. 0.70 (CI 0.61–0.78)<br>7. 0.73 (CI N/A)<br>8. 0.35 (CI 0.22–0.49)<br>9. 0.75 (CI 0.68–0.82)<br>10. 0.75 (CI N/A)<br>11. 0.64 (PP group) and 0.63 (ITD group)[a] (CI N/A) | 1. 68 (53)<br>2. 50 (38)<br>3. 60 (25)<br>4. 145 (87)<br>5. 251 (117)<br>6. 130 (70)<br>7. 37 (22)<br>8. 104 (72) with biopsy but only 94 with valid TE<br>9. 183 (94)<br>10. 98 (77)<br>11. 170 ITD[a] and 154 PP (103) | 1. NAS > 3 grade 1 for each of the three features<br>2. Presence of lobular inflammation and hepatocyte ballooning + NAS for grading<br>3. NAS ≥ 5<br>4. NAS ≥ 4 and at least one in ballooning ≥ 1 and inflammation ≥ 1<br>5. NAS ≥ 5<br>6. NAS ≥ 3; > 5% of hepatic steatosis and inflammation were seen with hepatocyte ballooning<br>7. SAF<br>8. NAS ≥ 3<br>9. Presence of steatosis and inflammation with ballooning<br>10. Hepatic steatosis and inflammation with hepatocyte injury<br>11. SAF | 1. 0.82 se and 0.73 spec cut-off > 5.3 kPa<br>2. N/A<br>3. 0.48 (0.28–0.69) se and 0.89 (0.73–0.97) spec<br>4. 0.44 (0.24–0.65) se and 0.97 (0.85–1.0) spec<br>5. Se 0.78 and spec 0.63 cut-off 7.7 kPa<br>6. N/A<br>7. 0.64 se and 0.87 spec at 9.9 kpa cut-off<br>8. 0.61 se and 0.59 spec cut-off of 5.6 kPa<br>9. Se 0.86 and spec 0.58 cut-off > 7 kPa<br>10. Se 0.73 and 0.75 spec cut-off 5.75 kPa<br>11. 0.54 se and 0.46 spec cut-off 6 kPa (PP[a]) and 0.54 se and 0.46 spec cut-off of 6 kPa (ITD group)[a] |
| Fibroscan + AST index | Kim et al. 2022 [47] | 0.74 (0.61–0.85) | 60 (25) | NAS ≥ 5 | 0.44 (0.24–0.65) se and 0.97 (0.85–1.0) spec |
| VCTE-LS (fibroscan) and CAP | Imajo et al. 2021 [43] | 0.71 (CI 0.66–0.8) | 145 (87) | NASH was defined as NAS ≥ 4 and at least one in ballooning ≥ 1 and inflammation ≥ 1 | N/A<br>No cut-off given |
| CLA model – CAP values greater than 250 dB/m, LS > 7 kPa, ALT level > 60 IU/L, hypertension, current smoker and presence of metabolic syndrome | Lee et al. 2016 [51] | 0.81 (CI 0.72–0.88)<br>The bootstrap method, average value of 0.83 (CI 0.74–0.89) | 183 (94) | NASH was defined as the presence of steatosis and inflammation with ballooning regardless of fibrosis | 1. Model N/A |

(Continues)

TABLE 4 | (Continued)

| Method/model used | Authors and year | AUROC (95% CL) | Population number (MASH) | Path grading | Sensitivity and specificity |
|---|---|---|---|---|---|
| FAST score | 1. Lee et al. 2022 [52] 2. Eilenberg et al. 2021 [32] per protocol (PP) analysis, excluding unreliable results as well as an intent-to-diagnose (ITD) analysis including failed measurements | 1. 0.75 (CI 0.69–0.81) 2. 0.78 (CI N/A) | 1. 251 (117) 2. 170 ITP and 154 PP (103) | 1. NAS ≥5 2. SAF | 1. Se 0.80 and spec 0.58 Fast cut-off 0.48 2. 0.59 se and 0.90 spec with a cut-off 366 (PP group) and se 0.62 and 0.87 spec (ITD group) |
| TE and BMI score | Salvatia et al. 2021 [66] | 0.82 (CI N/A) | 98 (77) | NASH was defined as hepatic steatosis and inflammation with hepatocyte injury (e.g., ballooning) | 0.75 se and 0.80 spec cut-off of 154.56 kPa•kg/m² 0.83 se and 0.83 spec with the cut-off of 98.11 kPa•kg/m² |
| Bayesian approach – CK-18 levels, hydrogen breath test and TE | Yilmaz et al. 2014 [78] | N/A | 235 (135) | NAS ≥5 | N/A |
| *CAP* | | | | | |
| CAP | 1. Garteiser et al. 2021 [38] 2. Imajo et al. 2021 [43] 3. Trowell et al. 2021 [75] 4. Troelstra et al. 2021 [76] 5. Lee et al. 2016 [51] 6. Eilenberg et al. 2021 [32] | 1. 0.69 (CI 0.59–0.79) 2. 0.71 (CI 0.62–0.80) 3. 0.80 (CI N/A) Validation 4. 0.65 (CI N/A) 5. 0.74 (CI 0.67–0.82) 6. 0.76 (PP group)[a] and 0.76 (ITD group)[a] (CI N/A) | 1. 152 (30) 2. 145 (87) 3. 217 (68) total cohort (Training 151 and validation 66) 4. 37 (22) 5. 183 (94) 6. 170 ITP and 154 PP (103) | 1. FLIP 2. NAS ≥4 and ballooning ≥1 and inflammation ≥1 3. NAS ≥3 4. Presence of steatosis where hepatocellular ballooning grade was ≥1 and where lobular inflammation grade ≥1 was present 5. Presence of steatosis and inflammation with ballooning regardless of fibrosis 6. SAF | 1. Se of 0.86 (0.71–0.95) and Sp of 0.47 (0.35–0.59), at a cut-off of 318 dB/m 2. N/A 3. Validation 0.82 se and 0.80 spec cut-off 301 dB/m 4. 0.77 se and 0.53 spec cut-off 336 dB/m 5. CAP >250 dB/m se 0.96 and spec 0.49 6. 0.80 se and 0.62 spec at cut-off of 323.5 dB/m+0.81 se and 0.60 spec at cut-off of 323.5 dB/m |

(Continues)

TABLE 4 | (Continued)

| Method/model used | Authors and year | AUROC (95% CL) | Population number (MASH) | Path grading | Sensitivity and specificity |
|---|---|---|---|---|---|
| CAST – CAP + AST | Imajo et al. 2023 [42] | 0.77 (CI 0.67–0.85) Validation cohort<br>0.69 (CI 0.62–0.76) In the US cohort | 169 (98) validation cohort<br>234 (117) USCD cohort | NAS ≥ 4 | N/A |
| CIR score – CAP, HOMA and insulin resistance | Macias et al. 2022 [59] | 0.93 (CI 0.87–0.99) Testing cohort | 155 (28) | SAF | N/A |

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; CAP, controlled attenuated parameter; CI, confidence interval; CK-18, cytokeratin 18; dB/m, decibels per meter; FAST, Fibroscan and AST; HOMA, homeostatic model for assessment of insulin resistance; ITD, intent-to-diagnose; kg/m², kilograms per metre squared; kPa, kilopascals; LS, liver stiffness; NAS, NAFLD activity score; PP, per protocol; SAF, steatosis activity and fibrosis score; Se, sensitivity; Spec, specificity; TE, transient elastography.
[a]Per protocol (PP) analysis excluding unreliable results and intent-to-diagnose (ITD) analysis based includes failed measurements.

method to give a more accurate perspective of available tests. We found that ultrasound and MRI are the most investigated imaging modalities and showed the most promise in our results for the diagnosis of MASH. Other modalities, including CT, TE, CAP and scintigraphy, had more variable results with limited studies and, in the case of TE, yielded better results when combined with other imaging modality. Ultrasound and MRI carry no radiation, unlike CT. Compared to MRI, ultrasound has a lower cost and can be used in patients with claustrophobia and metal implants/pacemakers that are not MRI-safe. However, adequate ultrasound imaging is restricted in obese patients.

Within the MRI studies, MRE studies especially showed good AUROCs for distinguishing between NASH and simple steatosis. However, Imajo et al. reported an AUROC of only 0.57 (CI: 0.47–0.67) in their study [43]. The highest AUROC of 0.93 was by Chen et al. Although the study did show that there was higher liver stiffness in patients with NASH without fibrosis over patients with simple steatosis, there was the most liver stiffness in those with fibrosis. The study used the Brunt criteria, which include fibrosis within its pathological definition, which may account for the higher AUROC [23]. This would suggest that, like any elastography method due to liver stiffness, MRE may yield better results in patients with MASH and fibrosis than in patients with MASH without fibrosis. However, other studies assessing MRE for identifying cases with MASH and fibrosis ≥ 2 suggest lower results, with an AUROC of only 0.66 in the study by Imajo et al. [43], and Costa-silva et al. report an AUROC of 0.79 for MRE using NAS ≥ 5, a NASH definition that does not include fibrosis [27]. SPIO has good results with AUROCs of 0.79 and 0.87, but there are safety concerns regarding this that limit its use [68, 73]. Two studies reported on the use of MRS, but the ability of MRS, high cost and long length of scans limit its use in clinical practice [20, 48]. Kim et al. found NASH patients had higher levels of Alanine (Ala) ($p < 0.001$) and Lactate (Lac) + Triglyceride (TG) ($p < 0.001$) than simple steatosis or controls. They found that Ala concentration had the best sensitivity of 100% and a specificity of 100% with a cut-off of > 16.04% (AUROC 1.0) [48]. The second larger study on MRS showed a lower 0.71 AUROC [20]. These studies did have different study designs. Abrigo et al. looked at biochemical changes with 31-Phosphate-MRS ($^{31}$P-MRS) and found that NASH patients had a decreased alpha nucleotide triphosphate (α-NTP)/total phosphate (TP) ($p < 0.001$) ratio. An α-NTP/TP cut-off of 16.36% gave 91% sensitivity, and a cut-off of 10.57% gave 91% specificity for NASH [20]. The Abrigo et al. study contained 151 patients (95 NASH) compared to 32 patients (11 NASH) in the Kim et al. study. The studies also used different pathology definitions [20, 48]. When evaluating MRI alone, Leporq et al. reported on a decrease in magnetic susceptibility in NASH leading to an AUROC of 0.91 [55] but we found no other studies on this technique to check repeatability. Additionally, Kim et al. report on a multiparametric MRI index obtaining an AUROC of 0.88, but again, [82] we found no other study on this model.

AUROCs > 0.90 were seen in a total of 7/27 ultrasound-based studies, [35, 39, 44, 50, 72, 81, 84] although one was for SW in a group of cirrhotic NASH patients [81]. This contrasts with MRI; despite more studies within this review, only two had an

**TABLE 5** | Diagnostic accuracy of other imaging methods in MASH diagnosis.

| Method/model used | Authors and year | AUROC (95% CL) | Population number (MASH) | Path grading | Sensitivity and specificity |
|---|---|---|---|---|---|
| **CT** | | | | | |
| 1. Entropy<br>2. MPP | Dichtel et al. 2023 [29] | 1. 0.88 (CI 0.62–0.76)<br>2. 0.90 (CI 0.62–0.76) | 16 (10) | NAS ≥1 in each category | 1. 0.8 se (CI 0.50–1.00) and 1.0 spec 95% CI: 67%–100%) entropy values below 4.73<br>2. 1.0 (0.50–1.00) se and 1.0 (67%–100%) spec MPP values below 54.0 |
| 1. Liver-spleen HU<br>2. Liver HU | Kim et al. 2023 [46] | 1. 0.82 L–S HU distinguishing NAS >4 versus <4 (CI N/A) + 0.75 for NAS <3 versus NAS ≥3<br>2. 0.86 (CI N/A) Liver HU NAS >4 versus <4 + 0.70 (CI N/A) for NAS <3 versus NAS ≥3 In patient group with low grade fatty liver disease | 142 (6) | NAS >4<br>G1 = Group NAS <3;<br>G2 = Group NAS 3–4<br>G3 = Group NAS ≥4 | 1. 0.71 se and 0.71 spec cut-off 3 for NAS ≥3 and L–S HU value, with a cut-off value of −1, NAS ≥5 with 1.0 se and 0.71 spec<br>2. N/A |
| CT parameters | Lubner et al. 2023 [58] | <0.60 for all individual CT features and subjective assessments | 186 (87) | Patients were considered to have NASH if hepatic steatosis and any lobular inflammation or hepatic ballooning was identified | N/A |
| Non-contrast enhanced CT<br>1. NASH prediction model for patients without suspicion of fibrosis – 12 parameters<br>2. NASH prediction model for patients with a suspicion of fibrosis using 2 parameters<br>2. Fibrosis was prediction by elevated levels of serum hyaluronic acid | Naganawa et al. 2018 [61] | 1. 0.94 (CI N/A) In patients without suspected fibrosis (texture parameter mean without filter and skewness with a 2-mm filter)<br>2. 0.60 (CI N/A) in patients with suspicion of fibrosis (the mean without filtration and kurtosis with a 4-mm filter) Validation cohort | 35 (NASH number N/A) validation | NAS was ≥3 | 1. 1.0 se and 0.92 spec<br>2. 1.0 se and 0.31 spec |

(Continues)

TABLE 5 | (Continued)

| Method/model used | Authors and year | AUROC (95% CL) | Population number (MASH) | Path grading | Sensitivity and specificity |
|---|---|---|---|---|---|
| CT parameters | Saadeh et al. 2002 [70] | N/A. None of the radiological features could distinguish patients with the pathologic diagnosis of NASH from steatosis | 25 (17) | Type 3 (steatosis, lobular inflammation, hepatocyte ballooning) and type 4 (steatosis, lobular inflammation, hepatocyte ballooning and Mallory's hyaline or fibrosis). Patients with types 3 and 4 were categorised as NASH | N/A |
| *Scintigraphy* | | | | | |
| 1. Scintigraphy 2. 9mTc-phytate scintigraphy | 1. D'Avila da Silva et al. 2021 [67] 2. Kikuchi et al. 2009 [45] | 1. N/A; the results did not reach the desired levels 2. 0.82 (CI N/A) | 1. 1.61 (49) 2. 37 (29) | 1. Presence of steatosis associated with ballooning and inflammatory infiltrate 2. NAS ≥ 5 | 1. N/A 2. Se of 1.0 and 0.75 spec with a cut-off point of 2.93, the liver/spleen uptake ratio |

Abbreviations: CI, Confidence Interval; L–S HU, liver spleen Hounsfield unit; MPP, mean value of positive pixels; NAS, NAFLD activity score; Se, sensitivity; Spec, specificity.

AUROC > 0.90 from our papers [28, 55]. Four of these studies contained SW in its model [35, 44, 50, 81]. SW itself had variable results ranging from 0.58 to 0.95, but the latter study was in cirrhotic NASH. This study achieved an AUROC of 0.88 for non-cirrhotic NASH in 2D-SWE and 0.95 in cirrhotic NASH, demonstrating elastography achieving better outcomes with more advanced fibrosis [81]. US based combination scores yielded the best results, such as the NASH pentagon by Funada et al. combining ultrasound parameters. The NASH pentagon achieved an AUROC of 0.94; the study did include fibrosis in their pathological diagnosis but did not report on any sensitivity and specificity levels [35]. However, Jang et al., with their study, also achieved an AUROC of 0.94 for a US risk score. They used the FLIP criteria. They also report good sensitivity (0.81) and specificity (0.96) with this study [44]. Three studies reported on US-FLI; this utilises five ultrasound parameters to give a diagnosis of NASH [25, 69, 84]. Nelson et al. achieved a high AUROC for differentiating NASH versus no NASH, but when comparing steatosis versus NASH, the AUROC was 0.65 (95% CI, 0 0.56–0.72) [84]. Ballestri et al. demonstrated a high AUROC of 0.80 when NAS ≥ 5 and a US-FLI score of 4 or lower suggested the absence of NASH (negative predictive value, 88%; sensitivity, 91% but specificity was 46%) [25]. The third study using POCUS and US-FLI found an AUROC of 0.78 and when US-FLI ≤ 3 and a sensitivity of 1.0 for ruling out NASH. At a cut-off ≥ 6, US-FLI had 63% specificity in the study. They used the NAS score but did not comment on the actual number used [69]. Overall, the studies suggest the score is better for ruling out NASH rather than differentiating it from steatosis. One of the disadvantages with this score is that one of the parameters is gallbladder blurring; therefore, the patient must have a gallbladder to evaluate. We found only two studies on ARFI that showed good AUROC, sensitivities and specificities [33, 39]. More studies on this technique are needed.

The studies unfortunately contained major sources of heterogenicity, which precluded a meta-analysis. Firstly, like other similar systematic reviews on diagnosis of MASH, we found our studies used heterogenous definitions of MASH, which makes comparisons between tests difficult. The definition used may cause better outcomes. For example, if fibrosis was included in the definition as with Brunt criteria, which can diagnose MASH based on the fibrosis stage alone, then elastography methods may have an advantage here. As our results illustrate, the best AUROC in MRE and 2D-SWE are in studies using the Brunt criteria [28, 81]. The high scores may reflect the diagnosis of fibrosis rather than the inflammatory component. These would likely create different outcomes if different pathological definitions were used. Secondly, imaging techniques were often different even within the same modularity, especially ultrasound, which utilised several different combinations of scores. Thirdly, different thresholds were used. Studies often used the Youden index to obtain the optional threshold for their study, with varying results between papers. For the 11 studies that looked at transient elastography, 7 documented the thresholds used within the text, and each of these studies used a different threshold [23, 32, 51, 52, 63, 66, 76]. Within the 11 papers that reported on 2D-MRE, 6 commented on the threshold used and only two studies used the same threshold [23, 28]. Additionally, 5/11 MRE studies did not report sensitivity or specificity

data [22, 43, 54, 56, 87], and there was only a maximum of two studies using the same defined pathology score [27, 63, 87, 88]. Thus, these variations precluded a meta-analysis.

Our study found that the most promising imaging tools are MRI techniques or ultrasound-based scores and confirmed there is potential for MASH diagnosis. Within the US studies, there are specific US parameters that have been found to be significant when diagnosing MASH, such as hepatic vessel blurring, attenuation coefficient and hepato-renal index. Further studies could focus on these identified parameters. Additionally, for future studies, there needs to be a standard pathological definition of MASH used to compare studies or the availability of digital histology images to allow re-scoring of biopsies. Furthermore, more validation studies are needed, which will help minimise the bias found in studies and prove the repeatability of the tests/scores. Many of these papers have developed a model that has not been tested elsewhere. Validating in larger studies can be difficult to undertake due to costs, time and availability of participants or data. Work is required in this area to make it easier for researchers to validate their studies. A new system, perhaps a large, easily accessible data pool, may help this. Without larger validation studies, there is no clear imaging tool or score currently available to diagnose MASH reliably.

## 5.1 | Study Limitations

Firstly, many of the studies suffered from bias. Introduction of a control group in the results, as seen in three of the papers, [20, 41, 44] may affect the accuracy of tests and does not reflect real-world testing of at-risk groups. Only 14 studies had pre-set cut-offs or offered a validation for their index test; thus, the reminders were deemed at high risk of bias for the index test section within the tool [34, 36, 37, 42, 53, 59–61, 69, 70, 75, 81, 86, 88]. This is understandable when assessing a new diagnostic test but will cause bias, particularly when cut-offs are adjusted to obtain higher sensitivity and specificity levels. For this systematic review, a comprehensive search was conducted, but studies without a clear report of liver biopsy or MASH diagnosis could have been omitted. Secondly, we did not analyse our planned secondary outcome of reviewing papers in relation to the components of the NAS score. This was only available in a limited selection of papers, and therefore felt not as valuable to the overall focus of the paper. However, this could be considered for future work.

## 6 | Conclusion

With advances in the treatment of MASH and studies showing MASH resolution resolves fibrosis and improves survival, then a focus on diagnosis of MASH rather than fibrosis is important. If we focus on fibrosis, then we would be missing an important group of patients with MASH and little/no fibrosis that could go on to develop fibrosis in the future and are currently subject to increased risk of cancer and lower survival than the general population. Imaging tools have been studied in the diagnosis of MASH, and there has been some progress made, especially ultrasound and MRI techniques. However, larger validation studies and an easier pathway for validation are needed.

## References

1. M. Rinella, J. Lazarus, V. Ratziu, et al., "A Multi-Society Delphi Consensus Statement on New Fatty Liver Disease Nomenclature," *Hepatology* 29, no. 1 (2023): 1542.

2. Z. M. Younossi, P. Golabi, J. M. Paik, A. Henry, C. Van Dongen, and L. Henry, "The Global Epidemiology of Nonalcoholic Fatty Liver Disease (NAFLD) and Nonalcoholic Steatohepatitis (NASH): A Systematic Review," *Hepatology* 77, no. 4 (2023): 1335–1347.

3. C. Estes, H. Razavi, R. Loomba, Z. Younossi, and A. J. Sanyal, "Modeling the Epidemic of Nonalcoholic Fatty Liver Disease Demonstrates an Exponential Increase in Burden of Disease," *Hepatology* 67, no. 1 (2018): 123–133.

4. H. Angus and N. James, "The Evolution of the Liver Transplant Candidate," *Frontiers in Transplantation* 2 (2023): 1–10.

5. Z. M. Younossi, M. Stepanova, Y. Younossi, et al., "Epidemiology of Chronic Liver Diseases in the USA in the Past Three Decades," *Gut* 69, no. 3 (2020): 564–568.

6. Z. M. Younossi, D. Blissett, R. Blissett, et al., "The Economic and Clinical Burden of Nonalcoholic Fatty Liver Disease in the United States and Europe," *Hepatology* 64, no. 5 (2016): 1577–1586.

7. A. Canbay, N. Kachru, J. S. Haas, D. Meise, A. B. Ozbay, and J. P. Sowa, "Healthcare Resource Utilization and Costs Among Nonalcoholic Fatty Liver Disease Patients in Germany," *Annals of Translational Medicine* 9, no. 8 (2021): 615.

8. P. S. Dulai, S. Singh, J. Patel, et al., "Increased Risk of Mortality by Fibrosis Stage in Nonalcoholic Fatty Liver Disease: Systematic Review and Meta-Analysis," *Hepatology* 65, no. 5 (2017): 1557–1565.

9. "FDA Approves Madrigal's Rezdiffra as the First NASH Therapy [Press Release], Online: Pharmaceutical Technology," (2024).

10. E. Vilar-Gomez, Y. Martinez-Perez, L. Calzadilla-Bertot, et al., "Weight Loss Through Lifestyle Modification Significantly Reduces

Features of Nonalcoholic Steatohepatitis," *Gastroenterology* 149, no. 2 (2015): 367–378.e5. quiz e14–e15.

11. O. Verrastro, S. Panunzi, L. Castagneto-Gissey, et al., "Bariatric-Metabolic Surgery Versus Lifestyle Intervention Plus Best Medical Care in Non-Alcoholic Steatohepatitis (BRAVES): A Multicentre, Open-Label, Randomised Trial," *Lancet* 401, no. 10390 (2023): 1786–1797.

12. S. Singh, A. M. Allen, Z. Wang, L. J. Prokop, M. H. Murad, and R. Loomba, "Fibrosis Progression in Nonalcoholic Fatty Liver vs Nonalcoholic Steatohepatitis: A Systematic Review and Meta-Analysis of Paired-Biopsy Studies," *Clinical Gastroenterology and Hepatology* 13, no. 4 (2015): 643–654.e9. quiz e39–40.

13. G. Lassailly, R. Caiazzo, V. Gnemmi, et al., *MASH Resolution Without Fibrosis Worsening After Bariatric Surgery Improves Long-Term Survival* Presented at: AASLD 2023; November 10-14. Boston, MA. Abstract 1. 2023.

14. J. Neuberger, J. Patel, H. Caldwell, et al., "Guidelines on the Use of Liver Biopsy in Clinical Practice From the British Society of Gastroenterology, the Royal College of Radiologists and the Royal College of Pathology," *Gut* 69, no. 8 (2020): 1382–1403.

15. G. Besutti, L. Valenti, G. Ligabue, et al., "Accuracy of Imaging Methods for Steatohepatitis Diagnosis in Non-Alcoholic Fatty Liver Disease Patients: A Systematic Review," *Liver International* 39, no. 8 (2019): 1521–1534.

16. SIGN, "Methodology Checklist 1: Systematic Reviews and Meta-Analyses," (2021), https://www.sign.ac.uk/what-we-do/methodology/checklists/.

17. J. J. B. P. Deeks, M. M. Leeflang, and Y. Takwoingi, *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*, 1st ed. (Chichester (UK): John Wiley & Sons, 2023).

18. PRISMA, "PRISMA for Diagnostic Test Accuracy Checklist," http://prisma-statement.org/Extensions/DTA.

19. F. Ravaioli, E. Dajti, A. Mantovani, P. N. Newsome, G. Targher, and A. Colecchia, "Diagnostic Accuracy of FibroScan-AST (FAST) Score for the Non-Invasive Identification of Patients With Fibrotic Non-Alcoholic Steatohepatitis: A Systematic Review and Meta-Analysis," *Gut* 72, no. 7 (2023): 1399–1409.

20. J. M. Abrigo, J. Shen, V. W. Wong, et al., "Non-Alcoholic Fatty Liver Disease: Spectral Patterns Observed From an In Vivo Phosphorus Magnetic Resonance Spectroscopy Study," *Journal of Hepatology* 60, no. 4 (2014): 809–815.

21. A. Al Manasra, M. Bani Hani, R. Khasawneh, et al., "Correlation Between Ultrasound and Histologic Findings of Fatty Liver Changes Among Non-Alcoholic Obese Patients," *Jordan Medical Journal* 55 (2021): 84–95.

22. A. M. Allen, V. H. Shah, T. M. Therneau, et al., "The Role of Three-Dimensional Magnetic Resonance Elastography in the Diagnosis of Nonalcoholic Steatohepatitis in Obese Patients Undergoing Bariatric Surgery," *Hepatology* 71, no. 2 (2020): 510–521.

23. S. Alsaqal, P. Hockings, H. Ahlström, et al., "The Combination of MR Elastography and Proton Density Fat Fraction Improves Diagnosis of Nonalcoholic Steatohepatitis," *Journal of Magnetic Resonance Imaging* 56, no. 2 (2022): 368–379.

24. V. B. Amorim, D. B. Parente, F. F. Paiva, et al., "Can Gadoxetic Acid-Enhanced Magnetic Resonance Imaging Be Used to Avoid Liver Biopsy in Patients With Nonalcoholic Fatty Liver Disease?," *World Journal of Hepatology* 12, no. 9 (2020): 661–671.

25. S. Ballestri, A. Lonardo, D. Romagnoli, et al., "Ultrasonographic Fatty Liver Indicator, a Novel Score Which Rules Out NASH and Is Correlated With Metabolic Parameters in NAFLD," *Liver International* 32, no. 8 (2012): 1242–1252.

26. N. Bastati, D. Feier, A. Wibmer, et al., "Noninvasive Differentiation of Simple Steatosis and Steatohepatitis by Using Gadoxetic Acid-Enhanced MR Imaging in Patients With Nonalcoholic Fatty Liver Disease: A Proof-Of-Concept Study," *Radiology* 271, no. 3 (2014): 739–747.

27. L. Costa-Silva, S. M. Ferolla, A. S. Lima, P. V. T. Vidigal, and T. C. A. Ferrari, "MR Elastography Is Effective for the Non-invasive Evaluation of Fibrosis and Necroinflammatory Activity in Patients With Nonalcoholic Fatty Liver Disease," *European Journal of Radiology* 98 (2018): 82–89.

28. J. Chen, J. A. Talwalkar, M. Yin, K. J. Glaser, S. O. Sanderson, and R. L. Ehman, "Early Detection of Nonalcoholic Steatohepatitis in Patients With Nonalcoholic Fatty Liver Disease by Using MR Elastography," *Radiology* 259, no. 3 (2011): 749–756.

29. L. E. Dichtel, A. Tabari, N. D. Mercaldo, et al., "CT Texture Analysis in Nonalcoholic Fatty Liver Disease (NAFLD)," *Journal of Clinical and Experimental Hepatology* 13, no. 5 (2023): 760–766.

30. B. Dzyubak, J. Li, J. Chen, et al., "Automated Analysis of Multiparametric Magnetic Resonance Imaging/Magnetic Resonance Elastography Exams for Prediction of Nonalcoholic Steatohepatitis," *Journal of Magnetic Resonance Imaging* 54, no. 1 (2021): 122–131.

31. P. J. Eddowes, N. McDonald, N. Davies, et al., "Utility and Cost Evaluation of Multiparametric Magnetic Resonance Imaging for the Assessment of Non-Alcoholic Fatty Liver Disease," *Alimentary Pharmacology & Therapeutics* 47, no. 5 (2018): 631–644.

32. M. Eilenberg, P. Munda, J. Stift, et al., "Accuracy of Non-invasive Liver Stiffness Measurement and Steatosis Quantification in Patients With Severe and Morbid Obesity," *Hepatobiliary Surgery and Nutrition* 10, no. 5 (2021): 610–622.

33. C. Fierbinteanu Braticevici, I. Sporea, E. Panaitescu, and L. Tribus, "Value of Acoustic Radiation Force Impulse Imaging Elastography for Non-invasive Evaluation of Patients With Nonalcoholic Fatty Liver Disease," *Ultrasound in Medicine and Biology* 39, no. 11 (2013): 1942–1950.

34. S. M. Francque, A. Verrijken, I. Mertens, et al., "Noninvasive Assessment of Nonalcoholic Fatty Liver Disease in Obese or Overweight Patients," *Clinical Gastroenterology and Hepatology* 10, no. 10 (2012): 1162–1168. quiz e87.

35. K. Funada, Y. Kusano, Y. Gyotoku, R. Shirahashi, T. Suda, and M. Tamano, "Novel Multi-Parametric Diagnosis of Non-Alcoholic Fatty Liver Disease Using Ultrasonography, Body Mass Index, and Fib-4 Index," *World Journal of Gastroenterology* 29, no. 23 (2023): 3703–3714.

36. R. Gallego-Durán, P. Cerro-Salido, E. Gomez-Gonzalez, et al., "Imaging Biomarkers for Steatohepatitis and Fibrosis Detection in Non-Alcoholic Fatty Liver Disease," *Scientific Reports* 6 (2016): 31421.

37. F. Gao, Q. He, G. Li, et al., "A Novel Quantitative Ultrasound Technique for Identifying Non-Alcoholic Steatohepatitis," *Liver International* 42, no. 1 (2022): 80–91.

38. P. Garteiser, L. Castera, M. Coupaye, et al., "Prospective Comparison of Transient Elastography, MRI and Serum Scores for Grading Steatosis and Detecting Non-Alcoholic Steatohepatitis in Bariatric Surgery Candidates," *JHEP Reports* 3, no. 6 (2021): 100381.

39. F. Guzmán-Aroca, M. D. Frutos-Bernal, A. Bas, et al., "Detection of Non-Alcoholic Steatohepatitis in Patients With Morbid Obesity Before Bariatric Surgery: Preliminary Evaluation With Acoustic Radiation Force Impulse Imaging," *European Radiology* 22, no. 11 (2012): 2525–2532.

40. I. S. Idilman, H. M. Low, T. Gidener, et al., "Association Between Visceral Adipose Tissue and Non-Alcoholic Steatohepatitis Histology in Patients With Known or Suspected Non-Alcoholic Fatty Liver Disease," *Journal of Clinical Medicine* 10, no. 12 (2021): 2565.

41. H. Iijima, F. Moriyasu, K. Tsuchiya, et al., "Decrease in Accumulation of Ultrasound Contrast Microbubbles in Non-Alcoholic Steatohepatitis," *Hepatology Research* 37, no. 9 (2007): 722–730.

42. K. Imajo, Y. Saigusa, T. Kobayashi, et al., "M-PAST Score Is Better Than MAST Score for the Diagnosis of Active Fibrotic Nonalcoholic Steatohepatitis," *Hepatology Research* 53, no. 9 (2023): 844–856.

43. K. Imajo, L. Tetlow, A. Dennis, et al., "Quantitative Multiparametric Magnetic Resonance Imaging Can Aid Non-Alcoholic Steatohepatitis Diagnosis in a Japanese Cohort," *World Journal of Gastroenterology* 27, no. 7 (2021): 609–623.

44. J. K. Jang, E. S. Lee, J. W. Seo, et al., "Two-Dimensional Shear-Wave Elastography and US Attenuation Imaging for Nonalcoholic Steatohepatitis Diagnosis: A Cross-Sectional, Multicenter Study," *Radiology* 305, no. 1 (2022): 118–126.

45. M. Kikuchi, K. Tomita, T. Nakahara, et al., "Utility of Quantitative 99mTc-Phytate Scintigraphy to Diagnose Early-Stage Non-Alcoholic Steatohepatitis," *Scandinavian Journal of Gastroenterology* 44, no. 2 (2009): 229–236.

46. H. N. Kim, H. J. Jeon, H. G. Choi, et al., "CT-Based Hounsfield Unit Values Reflect the Degree of Steatohepatitis in Patients With Low-Grade Fatty Liver Disease," *BMC Gastroenterology* 23, no. 1 (2023): 77.

47. J. W. Kim, C. H. Lee, B. H. Kim, et al., "Ultrasonographic Index for the Diagnosis of Non-Alcoholic Steatohepatitis in Patients With Non-Alcoholic Fatty Liver Disease," *Quantitative Imaging in Medicine and Surgery* 12, no. 3 (2022): 1815–1829.

48. T. H. Kim, H. Y. Jun, K. J. Kim, et al., "Hepatic Alanine Differentiates Nonalcoholic Steatohepatitis From Simple Steatosis in Humans and Mice: A Proton MR Spectroscopy Study With Long Echo Time," *Journal of Magnetic Resonance Imaging* 46, no. 5 (2017): 1298–1310.

49. Y. Kim, J. Hwang, S. H. Bae, et al., "Clinical Feasibility of Shear Wave Dispersion Slope for Noninvasive Diagnosis of Nonalcoholic Steatohepatitis in Patients With Morbid Obesity: Preliminary Results Using US Shear Wave Elastography," *Ultrasound Quarterly* 38, no. 2 (2022): 149–154.

50. D. H. Lee, E. J. Cho, J. S. Bae, et al., "Accuracy of Two-Dimensional Shear Wave Elastography and Attenuation Imaging for Evaluation of Patients With Nonalcoholic Steatohepatitis," *Clinical Gastroenterology and Hepatology* 19, no. 4 (2021): 797–805.e7.

51. H. W. Lee, S. Y. Park, S. U. Kim, et al., "Discrimination of Nonalcoholic Steatohepatitis Using Transient Elastography in Patients With Nonalcoholic Fatty Liver Disease," *PLoS One* 11, no. 6 (2016): e0157358.

52. J. S. Lee, B. K. Kim, J. Y. Park, et al., "Comparison of FibroScan-Aspartate Aminotransferase (FAST) Score and Other Non-Invasive Surrogates in Predicting High-Risk Non-Alcoholic Steatohepatitis Criteria," *Frontiers in Medicine* 9 (2022): 869190.

53. Y.-S. Lee, J. E. Lee, H.-S. Yi, et al., "MRE-Based NASH Score for Diagnosis of Nonalcoholic Steatohepatitis in Patients With Nonalcoholic Fatty Liver Disease," *Hepatology International* 16, no. 2 (2022): 316–324.

54. Y. S. Lee, Y. J. Yoo, Y. K. Jung, et al., "Multiparametric MR Is a Valuable Modality for Evaluating Disease Severity of Nonalcoholic Fatty Liver Disease," *Clinical and Translational Gastroenterology* 11, no. 4 (2020): e00157.

55. B. Leporq, S. A. Lambert, M. Ronot, V. Vilgrain, and B. E. Van Beers, "Simultaneous MR Quantification of Hepatic Fat Content, Fatty Acid Composition, Transverse Relaxation Time and Magnetic Susceptibility for the Diagnosis of Non-Alcoholic Steatohepatitis," *NMR in Biomedicine* 30, no. 10 (2017): e3766.

56. J. Li, X. Lu, Z. Zhu, et al., "Head-To-Head Comparison of Magnetic Resonance Elastography-Based Liver Stiffness, Fat Fraction, and T1 Relaxation Time in Identifying At-Risk NASH," *Hepatology* 78, no. 4 (2023): 1200–1208.

57. R. J. Liang, H. H. Wang, W. J. Lee, P. L. Liew, J. T. Lin, and M. S. Wu, "Diagnostic Value of Ultrasonographic Examination for Nonalcoholic Steatohepatitis in Morbidly Obese Patients Undergoing Laparoscopic Bariatric Surgery," *Obesity Surgery* 17, no. 1 (2007): 45–56.

58. M. G. Lubner, P. M. Graffy, A. Said, et al., "Utility of Multiparametric CT for Identification of High-Risk NAFLD," *AJR. American Journal of Roentgenology* 216, no. 3 (2021): 659–668.

59. J. Macias, P. Parra-Membrives, F. Sosa-Moreno, et al., "Controlled Attenuation Parameter-Insulin Resistance (CIR) Score to Predict Non-Alcoholic Steatohepatitis," *Scientific Reports* 12, no. 1 (2022): 21897.

60. D. Marti-Aguado, J. Arnouk, J. X. Liang, et al., "Development and Validation of an Image Biomarker to Identify Metabolic Dysfunction Associated Steatohepatitis: MR-MASH Score," *Liver International* 44, no. 1 (2023): 202–213.

61. S. Naganawa, K. Enooku, R. Tateishi, et al., "Imaging Prediction of Nonalcoholic Steatohepatitis Using Computed Tomography Texture Analysis," *European Radiology* 28, no. 7 (2018): 3050–3058.

62. A. Ozturk, R. Mohammadi, T. T. Pierce, et al., "Diagnostic Accuracy of Shear Wave Elastography as a Non-Invasive Biomarker of High-Risk Non-Alcoholic Steatohepatitis in Patients With Non-Alcoholic Fatty Liver Disease," *Ultrasound in Medicine & Biology* 46, no. 4 (2020): 972–980.

63. C. C. Park, P. Nguyen, C. Hernandez, et al., "Magnetic Resonance Elastography vs. Transient Elastography in Detection of Fibrosis and Noninvasive Measurement of Steatosis in Patients With Biopsy-Proven Nonalcoholic Fatty Liver Disease," *Journal of Hepatology* 66, no. 1 Supplement 1 (2017): S233.

64. M. Pavlides, R. Banerjee, E. M. Tunnicliffe, et al., "Multiparametric Magnetic Resonance Imaging for the Assessment of Non-Alcoholic Fatty Liver Disease Severity," *Liver International* 37, no. 7 (2017): 1065–1073.

65. F. B. Pulzi, R. Cisternas, M. R. Melo, C. M. Ribeiro, C. A. Malheiros, and J. E. Salles, "New Clinical Score to Diagnose Nonalcoholic Steatohepatitis in Obese Patients," *Diabetology and Metabolic Syndrome* 3, no. 1 (2011): 3.

66. A. Salvati, F. Faita, D. Cavallone, et al., "Steatosis/Steatohepatitis: How Sustainable Is the Non-Invasive Instrumental Differential Diagnosis in Clinical Practice?," *Hepatoma Research* 7 (2021): 14.

67. A. D. Da Silva, A. Z. de Mattos, C. V. Tovo, et al., "Nonalcoholic Fatty Liver Disease: Scintigraphy in the Diagnosis of Steatohepatitis," *Revista da Associação Médica Brasileira* 67, no. 11 (2021): 1665–1669.

68. L. P. Smits, B. F. Coolen, M. D. Panno, et al., "Noninvasive Differentiation Between Hepatic Steatosis and Steatohepatitis With MR Imaging Enhanced With USPIOs in Patients With Nonalcoholic Fatty Liver Disease: A Proof-Of-Concept Study," *Radiology* 278, no. 3 (2016): 782–791.

69. A. Sourianarayanane and A. J. McCullough, "Accuracy of Ultrasonographic Fatty Liver Index Using Point-Of-Care Ultrasound in Stratifying Non-Alcoholic Fatty Liver Disease Patients," *European Journal of Gastroenterology & Hepatology* 35, no. 6 (2023): 654–661.

70. S. Saadeh, Z. M. Younossi, E. M. Remer, et al., "The Utility of Radiological Imaging in Nonalcoholic Fatty Liver Disease," *Gastroenterology* 123, no. 3 (2002): 745–750.

71. K. Sugimoto, F. Moriyasu, H. Oshiro, et al., "The Role of Multiparametric US of the Liver for the Evaluation of Nonalcoholic Steatohepatitis," *Radiology* 296, no. 3 (2020): 532–540.

72. G. Tarantino, P. Conca, F. Pasanisi, et al., "Could Inflammatory Markers Help Diagnose Nonalcoholic Steatohepatitis?," *European Journal of Gastroenterology & Hepatology* 21, no. 5 (2009): 504–511.

73. K. Tomita, A. Tanimoto, R. Irie, et al., "Evaluating the Severity of Nonalcoholic Steatohepatitis With Superparamagnetic Iron Oxide-Enhanced Magnetic Resonance Imaging," *Journal of Magnetic Resonance Imaging* 28, no. 6 (2008): 1444–1450.

74. M. Torkzaban, C. E. Wessner, D. Halegoua-DeMarzio, A. Lyshchik, and K. Nam, "Diagnostic Performance of Quantitative Ultrasound

Parameters in Non-Alcoholic Fatty Liver Disease," *Academic Radiology* 30 (2023): 1838–1845.

75. J. Trowell, J. Alukal, T. Zhang, et al., "How Good Are Controlled Attenuation Parameter Scores From Fibroscan to Assess Steatosis, NASH, and Fibrosis?," *Digestive Diseases and Sciences* 66, no. 4 (2021): 1297–1305.

76. M. A. Troelstra, J. J. Witjes, A. M. van Dijk, et al., "Assessment of Imaging Modalities Against Liver Biopsy in Nonalcoholic Fatty Liver Disease: The Amsterdam NAFLD-NASH Cohort," *Journal of Magnetic Resonance Imaging* 54, no. 6 (2021): 1937–1949.

77. B. Wildman-Tobriner, M. M. Middleton, C. A. Moylan, et al., "Association Between Magnetic Resonance Imaging-Proton Density Fat Fraction and Liver Histology Features in Patients With Nonalcoholic Fatty Liver Disease or Nonalcoholic Steatohepatitis," *Gastroenterology* 155, no. 5 (2018): 1428–1435.e2.

78. Y. Yilmaz and F. Eren, "A Bayesian Approach to an Integrated Multimodal Noninvasive Diagnosis of Definitive Nonalcoholic Steatohepatitis in the Spectrum of Nonalcoholic Fatty Liver Disease," *European Journal of Gastroenterology & Hepatology* 26, no. 11 (2014): 1292–1295.

79. E. M. Zardi, I. De Sio, G. Ghittoni, et al., "Which Clinical and Sonographic Parameters May Be Useful to Discriminate NASH From Steatosis?," *Journal of Clinical Gastroenterology* 45, no. 1 (2011): 59–63.

80. Y. Zhao, C. Qiu, Y. Dong, et al., "Technical Acoustic Measurements Combined With Clinical Parameters for the Differential Diagnosis of Nonalcoholic Steatohepatitis," *Diagnostics (Basel)* 13, no. 9 (2023): 1547.

81. J. Zhou, F. Yan, J. Xu, et al., "Diagnosis of Steatohepatitis and Fibrosis in Biopsy-Proven Nonalcoholic Fatty Liver Diseases: Including Two-Dimension Real-Time Shear Wave Elastography and Noninvasive Fibrotic Biomarker Scores," *Quantitative Imaging in Medicine and Surgery* 12, no. 3 (2022): 1800–1814.

82. J. W. Kim, Y. S. Lee, Y. S. Park, et al., "Multiparametric MR Index for the Diagnosis of Non-Alcoholic Steatohepatitis in Patients With Non-Alcoholic Fatty Liver Disease," *Scientific Reports* 10, no. 1 (2020): 2671.

83. D. B. Parente, F. F. Paiva, J. A. Oliveira Neto, et al., "Intravoxel Incoherent Motion Diffusion Weighted MR Imaging at 3.0 T: Assessment of Steatohepatitis and Fibrosis Compared With Liver Biopsy in Type 2 Diabetic Patients," *PLoS One* 10, no. 5 (2015): e0125653.

84. S. M. Nelson, J. D. Hoskins, C. Lisanti, and J. Chaudhuri, "Ultrasound Fatty Liver Indicator: A Simple Tool for Differentiating Steatosis From Nonalcoholic Steatohepatitis: Validity in the Average Obese Population," *Journal of Ultrasound in Medicine* 39, no. 4 (2020): 749–759.

85. F. Lirussi, N. Vitturi, L. Azzalini, et al., "Perihepatic Adipose Tissue Thickness: A New Non-invasive Marker of NAFLD?," *Journal of Gastrointestinal and Liver Diseases* 18, no. 1 (2009): 61–66.

86. A. Petrick, P. Benotti, G. C. Wood, et al., "Utility of Ultrasound, Transaminases, and Visual Inspection to Assess Nonalcoholic Fatty Liver Disease in Bariatric Surgery Patients," *Obesity Surgery* 25, no. 12 (2015): 2368–2375.

87. R. Loomba, J. Cui, T. Wolfson, et al., "Novel 3D Magnetic Resonance Elastography for the Noninvasive Diagnosis of Advanced Fibrosis in NAFLD: A Prospective Study," *American Journal of Gastroenterology* 111, no. 7 (2016): 986–994.

88. R. Loomba, T. Wolfson, B. Ang, et al., "Magnetic Resonance Elastography Predicts Advanced Fibrosis in Patients With Nonalcoholic Fatty Liver Disease: A Prospective Study," *Hepatology* 60, no. 6 (2014): 1920–1928.

89. A. Santiago-Rolón, D. Purcell, K. Rosado, and D. H. Toro, "A Comparison of Brunt's Criteria, the Non-Alcoholic Fatty Liver Disease Activity Score (NAS), and a Proposed NAS Scoring That Includes Fibrosis in Non-Alcoholic Fatty Liver Disease Staging," *Puerto Rico Health Sciences Journal* 34, no. 4 (2015): 189–194.

90. P. Bedossa, "Utility and Appropriateness of the Fatty Liver Inhibition of Progression (FLIP) Algorithm and Steatosis, Activity, and Fibrosis (SAF) Score in the Evaluation of Biopsies of Nonalcoholic Fatty Liver Disease," *Hepatology* 60, no. 2 (2014): 565–575.

91. A. Guida, "NAFLD Study: US vs Liver Biopsy (NAFLD) Clinical," Trials.gov, https://clinicaltrials.gov/study/NCT04101162.

92. M. Dollinger, "Non-Invasive Rapid Assessment of NAFLD Using Magnetic Resonance Imaging With LiverMultiScan (RADIcAL1)," https://clinicaltrials.gov/study/NCT03289897.

93. University of California D and Diabetes NIo, Digestive, Diseases K, "Non-Alcoholic Fatty Liver Disease Parametric PET (FLiPP) Study," (2016), https://classic.clinicaltrials.gov/show/NCT02754037.

94. University TM, "Innovative Liver Elasticity, Attenuation, and Dispersion Ultrasound Study," (2019), https://classic.clinicaltrials.gov/show/NCT04012242.

95. Kong CUoH, "Simultaneous Quantification of Liver Fat Content, Fatty Acid Composition, and Fibrosis Using Spin-lock MRI for Steatohepatitis Assessment," (2024), https://classic.clinicaltrials.gov/show/NCT05384652.

96. Oxford Uo, University N, Nottingham Uo, et al., "LITMUS Imaging Study," (2019), https://classic.clinicaltrials.gov/show/NCT05479721.

97. S. M. Atzori, Y. Pasha, J. B. Maurice, S. D. Taylor-Robinson, L. Campbell, and A. K. P. Lim, "Prospective Evaluation of Liver Shear-wave Elastography Measurements With 3 Different Technologies and Same Day Liver Biopsy in Patients With Chronic Liver Disease," *Digestive and Liver Disease* 56, no. 3 (2024): 484–494.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section.