DR. GIULIA BESUTTI (Orcid ID : 0000-0001-5319-5495)

DR. LUCA VALENTI (Orcid ID : 0000-0001-8909-0345)

Article type : Original Articles

Accuracy of Imaging Methods for Steatohepatitis Diagnosis in Non-alcoholic Fatty Liver Disease Patients: A Systematic Review

Giulia Besutti [1,2], Luca Valenti [3], Guido Ligabue [4], Maria Chiara Bassi [5], Pierpaolo Pattacini [2], Giovanni Guaraldi [6], Paolo Giorgi Rossi [7].

1. Clinical and Experimental Medicine PhD Program, University of Modena and Reggio Emilia, Modena, Italy.

2. Radiology Unit, Azienda Unità Sanitaria Locale-IRCCS of Reggio Emilia, Reggio Emilia, Italy.

3. Department of Pathophysiology and Transplantation, Università degli Studi di Milano, and Translational Medicine, Department of Transfusion Medicine and Hematology, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy.

4. Radiology Unit, University of Modena and Reggio Emilia, Modena, Italy.

5. Medical Library, Azienda Unità Sanitaria Locale-IRCCS of Reggio Emilia, Reggio Emilia, Italy.

6. Modena HIV Metabolic Clinic, University of Modena and Reggio Emilia, Modena, Italy.

7. Epidemiology Unit, Azienda Unità Sanitaria Locale-IRCCS of Reggio Emilia, Reggio Emilia, Italy.

Corresponding author:

Giulia Besutti, MD

Clinical and Experimental Medicine PhD Program, University of Modena and Reggio Emilia, Modena, Italy.

Radiology Unit, Azienda Unità Sanitaria Locale-IRCCS of Reggio Emilia, Viale Risorgimento 80, 42123, Reggio Emilia, Italy. Telephone number: +39 0522 29 6393. E-mail: giulia.besutti@ausl.re.it.

Abbreviations:

NAFLD, non-alcoholic fatty liver disease

NASH, non-alcoholic steatohepatitis

MRS, magnetic resonance spectroscopy

PDFF, proton density fat fraction

AUROC, area under the receiving operating characteristic curve

QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies-2

TE, transient elastography

ARFI, acoustic radiation force impulse

MRE, magnetic resonance elastography

SPIO / USPIO, superparamagnetic iron oxide / ultrasmall superparamagnetic iron oxide

Conflicts of Interest: The authors have no conflicts of interest to declare.

Funding: The authors did not receive any financial support for this work.

ABSTRACT

Background & Aims: Non-invasive tests to diagnose non-alcoholic steatohepatitis (NASH) are urgently needed. This systematic review aims to evaluate imaging accuracy in diagnosing NASH among non-alcoholic fatty liver disease (NAFLD) patients, using liver biopsy as reference.

Methods: Eligible studies were systematic reviews and cross-sectional/cohort studies of NAFLD patients comparing imaging with histology, considering accuracy and/or associations. MEDLINE, Scopus, EMBASE, and Cochrane Library databases were searched up to April 2018. Studies were screened on title/abstract, then assessed for eligibility on full-text. Data were extracted using a pre-designed form. Risk of bias was assessed using Quality Assessment of Diagnostic Accuracy Studies-2 tool.

Results: Of the 641 studies screened, 58 were included in scoping review, 30 of which (with accuracy results) in data synthesis. Imaging techniques included: elastography (transient elastography-TE, acoustic radiation force impulse-ARFI, magnetic resonance elastography-MRE), ultrasound (US), magnetic resonance (MR), computed tomography and scintigraphy. Histological NASH definition was heterogeneous. In 28/30 studies, no prespecified threshold was used (high risk of bias). AUROCs were up to 0.82 for TE, 0.90 for ARFI, 0.93 for MRE and 0.82 for US scores. MR techniques with higher accuracy were spectroscopy (AUROC=1 for alanine), susceptibility-weighted imaging (AUROC=0.91), multiparametric MR (AUROC=0.80), optical analysis (AUROC=0.83), gadoxetic acid-enhanced MR (AUROCs=0.85) and superparamagnetic iron oxide-enhanced MR (AUROC=0.87). Results derived mostly from single studies without independent prospective validation.

Conclusions: There is currently insufficient evidence to support the use of imaging to diagnose NASH. More studies are needed on US and MR elastography and non-elastographic techniques, to date the most promising methods.

Keywords: Non-alcoholic Fatty Liver Disease; Non-alcoholic Steatohepatitis; Magnetic Resonance; Ultrasonography.

Abstract

Lay Summary: We identified several imaging techniques that were tested for accuracy in diagnosing steatohepatitis among patients with fatty liver, some with promising results (mostly ultrasound and magnetic resonance techniques). Studies were conducted on few patients, with different clinical features, using various definitions of steatohepatitis and without independent validation. Hence, more studies are needed on the most promising techniques.

INTRODUCTION

The estimated overall global prevalence of non-alcoholic fatty liver disease (NAFLD) is around 25% and projected at 33.5% in 2030.[1] While simple steatosis without evidence of inflammation and hepatocellular injury (non-alcoholic fatty liver) is generally a benign condition, non-alcoholic steatohepatitis (NASH) can progress to fibrosis, cirrhosis, liver failure and hepatocellular carcinoma.

Since only histological analysis can accurately evaluate NAFLD patterns, liver biopsy is the gold standard for assessment, and it should be considered in patients who are at increased risk of having steatohepatitis and/or fibrosis.[2] Major drawbacks are its invasive nature, risk of complications, sampling errors and inter and intra-observer variability.[3]

Currently, there are no approved therapies for NASH. However, several drugs are now in phase 2 and 3 trials, and results are expected in 1-2 years.[4] If medical treatments become available, screening for steatohepatitis and fibrosis will be recommended in high-risk patients. The lack of non-invasive tools to identify patients who may benefit from a

therapeutic intervention is a central issue. Should liver biopsy be avoided or reserved for a more limited number of undetermined or high-risk patients, the benefit-harm balance of NASH screening and therapies would undergo a major change.

Non-invasive imaging modalities such as magnetic resonance imaging (MRI) or spectroscopy (MRS) with calculation of proton density fat fraction (PDFF) accurately measure hepatic fat.[5] On the other hand, since fibrosis is the most important histological feature associated with long-term mortality in patients with NAFLD,[6] research on non-invasive tests, either serum biomarkers and imaging-based techniques, have focused on this outcome.[7,8]

However, the diagnosis of NASH provides important prognostic information indicating an increased risk of fibrosis progression, prompting a closer follow-up, and its resolution represents the main outcome for clinical trials.[9] Several marker panels have been proposed to differentiate between simple steatosis and NASH, with inconsistent results.[10] Some imaging methods, mostly ultrasound (US) or MR techniques, have shown promising potential in NASH diagnosis.

The objective of this systematic review is to evaluate the diagnostic accuracy of non-invasive imaging techniques in diagnosing NASH with or without fibrosis in patients with or at high risk of NAFLD, using liver biopsy as the reference standard.

MATERIALS AND METHODS

This review was conducted in two phases: 1) a scoping review aimed at mapping all the imaging tests proposed in the literature for NASH diagnosis; 2) data synthesis for those tests for which accuracy studies were available.

Study eligibility

Eligible studies were systematic reviews of studies comparing imaging and histology in the diagnosis of NASH and cross-sectional (prospective or retrospective) and cohort studies comparing one or more imaging techniques with the reference standard (liver histology). Complete protocol has been registered in the PROSPERO database (ID CRD42018089989). Only studies that recruited patients with an available direct NAFLD assessment (biopsy- or imaging-proven) or patients at high risk of NAFLD based on metabolic factors met the inclusion criteria.

Only studies considering the following outcomes were included: diagnostic accuracy in terms of sensitivity and specificity or area under the receiving operating characteristic curve (AUROC) (main outcome), associations between index test and reference standard and reproducibility (secondary outcomes).

Since the evaluation of the presence and resolution of NASH is currently the main goal of histological assessment of liver damage in patients with NAFLD, studies focusing only on the assessment of fibrosis or steatosis, without a specific aim at differentiation between simple steatosis and NASH, were not included.

Studies reported only as abstracts or published in languages other than English were excluded.

Study search and selection

A systematic search was conducted in MEDLINE, The Cochrane Library, EMBASE and Scopus, adapting the search algorithm to the requirement of each database. No limit was applied in terms of publication date. References of included studies were reviewed to identify any additional relevant study. The last search was conducted in April 2018. The search algorithm designed for MEDLINE is reported in Supplementary Methods section.

One reviewer (GB) screened the search results based on title/abstract; a second reviewer (PGR) screened a computer-generated random sample of 20% of the references to identify potential sources of disagreement, which were resolved by consensus. Then, one reviewer (GB) examined eligibility based on the full text of the relevant articles. When unclear, inclusion was decided by group consensus. Reasons for exclusion are reported in Supplementary Table 1.

Data extraction and synthesis

One reviewer (GB) extracted data on study design, country, objective, population (number and characteristics of included patients), technical information on imaging techniques, histological classification system, outcomes, prevalence of steatohepatitis and results. These data were collected in a pre-designed data extraction sheet. A cross-check of the extracted data for accuracy was conducted by another reviewer (PGR). The Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool[11] was used by two reviewers (GB, PGR) to assess the risk of bias by consensus.

Summary statistics were used to describe the studies, subjects and outcomes. Data pooling would be considered only for sensitivity and specificity, and in case of sufficient homogeneity of outcomes, diagnostic techniques and procedures. Furthermore, data reporting would be necessary to allow the use of consistent positivity thresholds when needed. Otherwise, only narrative synthesis would be done. The quality of the evidence was rated with the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.[12] Test-related consequences were considered only for those techniques with contrast media or radiotracer administration, or radiation exposure. Resource consumption in terms of human and technological resources, operator-dependence, and stage of development

according to the presence of harmonized procedures and defined/agreed positivity thresholds were also taken into consideration.

RESULTS

Characteristics of the included studies

Study selection according to the PRISMA flow diagram[13] is reported in Figure 1. Sixty-one studies met eligibility criteria for scoping review; of these, 30 reported accuracy results. Included studies were carried out from 1999 to 2018, principally in Europe, the United States and Japan. No systematic review specifically addressing imaging test for NASH diagnosis was found. Fourteen studies were retrospective, 46 were prospective, and one was described as mixed retrospective/prospective.

Population

All studies included patients with proven NAFLD or at high risk of NAFLD and NASH. The number of patients ranged from 8 to 513, with a total of 4693 patients included, though the number of tested patients for each technique was much smaller. Eighteen studies included a control group of healthy subjects, tested with index test but not with liver biopsy. These patients were not considered for accuracy measures in this review. Five studies also considered a subgroup of patients affected with a chronic liver disease other than NAFLD; in none of these studies was an accuracy analysis performed. Two studies were specifically conducted on children or young adults, two on patients with type 2 diabetes and 4 on morbidly obese bariatric surgery patients. Most patients in the remaining studies were also overweight or obese, with mean body mass index ranging from 25 to 38.

Index Tests

Most studies compared a single imaging technique with histology; seventeen evaluated and compared more than one technique (Figure 2). Figure 3 classifies the index tests in a matrix of the types of imaging techniques and the targeted physical feature.

Histopathological analysis

Liver histology was mostly obtained through US-guided percutaneous biopsy; in 8 studies it was obtained from intra-operative biopsies or surgical specimens.

Heterogeneous histopathological definitions of NASH were used.[14-17] The accepted definition of NASH as the contemporary presence of steatosis, lobular inflammation and ballooning independently of fibrosis was generally followed, but in 6 studies fibrosis was included in the definition of NASH or classified with NASH.[18-23] Even among studies which referred to the most used classification by the Clinical Research Network,[17] cases defined as borderline or with NAFLD Activity Score 3-4 were either classified with simple steatosis or with NASH. NASH prevalence ranged from 32% to 90%.

Outcomes

The main outcome (diagnostic accuracy for NASH diagnosis, i.e. differentiation between simple steatosis and NASH) was considered in 30 studies. In 4 of these, accuracy was measured in terms of AUROC, without identification of a cut-off value, while in the other 26 optimal cut-off values were reported with respective sensitivities and specificities. The remaining 31 studies reported only associations between index test and histopathological assessment (Supplementary Table 2). Reproducibility was only evaluated in a minority of the included studies (n=9).

Risk of bias analysis

Results are reported in Supplementary Figure 2 and Supplementary Tables 3 and 4. Apart from two studies which included an estimation cohort and a validation cohort,[24,25] all studies were judged at high risk of bias introduced by the index test because no prespecified thresholds were used. Patient selection introduced a high risk of bias in nearly 50% of the accuracy studies.

Synthesis of accuracy results

Because of the large heterogeneity in imaging techniques and technical parameters, positivity thresholds, and NASH histopathologic definition, data pooling was not possible. In this narrative synthesis (Table 1-4), only the 30 studies reporting accuracy are considered. A more detailed description of accuracy results and a synthesis of secondary outcomes are reported in Supplementary Results section and Supplementary Tables 5-8. The level of the certainty of the evidence, according to GRADE criteria, is reported for each technique in supplementary Table 9, and results are summarized in Table 5.

Among elastographic techniques (Table 1), the accuracy of TE was evaluated in four studies with different histopathologic definitions of NASH, showing AUROCs ranging from 0.65 (0.54-0.77) to 0.75 (0.68-0.82) for definite NASH, with sensitivity/specificity up to 86%/58% for NAS≥5 and 89%/90% for high-risk patients (NASH or fibrosis>1).[18,26,27] ARFI was evaluated in two studies, both with high risk of bias, resulting in sensitivities of 77%-85% and specificities of 72%-83%, using similar cut-off values.[19,28] MRE was evaluated for NASH diagnosis in six studies, again with different NASH definitions, resulting in AUROCs ranging from 0.70 to 0.79 in studies not including fibrosis in NASH definition,[5,18,29-31] with

sensitivity and specificity of 72% and 87% for NAS≥5 and similar results in a subset of patients without fibrosis.[29]

US non-elastographic techniques (Table 2) include several parameters and scores that took into consideration features related to the severity of steatosis, spleen diameter or visceral adiposity, all evaluated in one single study,[20,22,32-35] resulting in AUROCs ranging from 0.76 of US-fatty liver indicator (US-FLI) for NAS≥2 to 0.92 of splenic diameter. With a cut-off of 4, US-FLI presented 100% sensitivity and 46% specificity for the diagnosis of severe NASH.[32] The accuracy of contrast-enhanced US for NASH diagnosis was evaluated in one single study limited by partial verification, with sensitivity and specificity up to 100%.[36] Among MR non-elastographic techniques (Table 3), the ^{31}P-MRS-derived ratio between nucleotide triphosphates (α-peak) and triphosphates (αNTP/TP), reflecting cellular energetic failure,[21] and the concentration of specific metabolites (e.g. alanine, lactate, triglycerides) assessed by ^{1}H-MRS,[37] showed AUROCs ranging from 0.71 for αNTP/TP and 1.00 for alanine, the latter evaluated in a small sample of 26 patients for NAS ≥5. Multiparametric MRI (mpMRI) demonstrated AUROCs of 0.69, 0.74 and 0.80, respectively, in the differentiation between NASH and SS when considering corrected T1 (cT1) as index test, in the differentiation between NAS<5 and ≥5 for the same index test, and in the diagnosis of NASH by using Liver Inflammation and Fibrosis (LIF) score.[26,38] An optimal cut-off for LIF has recently been identified (1.4), with sensitivity 91% and specificity 52%. For cT1 as well, an optimal cut-off (875 ms) has been suggested, but to distinguish between low- and high-risk (NASH or fibrosis>1) patients, with sensitivity/specificity of 97%/50%. Other MRI approaches include quantitative susceptibility imaging,[39] intravoxel incoherent motion (IVIM) diffusion-weighted MRI,[40] and morphological evaluation such as liver volume [41] and preperitoneal fat area [42], all evaluated in one single study, with AUROCs ranging from

0.61/0.68/0.74 for different IVIM parameters to 0.91 for susceptibility, the last one tested in a small sample of 32 patients. Moreover, a score based on MRI optical analysis estimators produced an AUROC of 0.83 with sensitivity/specificity of 87%/60%.[24] Concerning contrast media-based approaches, gadoxetic acid enhancement in hepatobiliary phase showed sensitivity/specificity of 97%/63% in a retrospective study of 81 patients,[43] while superparamagnetic iron oxide (SPIO) and ultrasmall SPIO (USPIO)-enhanced MRI-derived ΔR2* demonstrated sensitivity/specificity up to 91%/73% for USPIO in a study of 25 patients for NAS≥5.[44,45]

Among other techniques (Table 4), CT texture features and TC99m-phytate colloid scintigraphy were assessed in small series (n=35 and 37 patients), resulting in AUROCs up to 0.94 and 0.82, respectively.[25,46]

The presence of direct consequence of the test on the health, the qualitative analysis of resource consumption, operator-dependence, and the state of the art of the techniques are reported in Table 5.

DISCUSSION

We found more than 40 different tests proposed for non-invasive diagnosis of NASH. Tests were based on at least four different principles, including quantification of liver stiffness, anatomical features, tissue composition and functional features, combined with four imaging modalities: ultrasound, MR, CT and scintigraphy. Several authors proposed scores based on combinations of different characteristics usually collected through the same imaging approach. This landscape produced an enormous quantity of possible tests, each one proposed by one or few groups of researchers but lacking robust and independent validation. Although the first study retrieved was from 1999, indicating almost 20 years of research in the field, the picture remains that of an early stage of development of the putative technologies. Indeed,

when more than one study was present, procedures and positivity thresholds were not uniform, and pooling of results was not possible. Another sign of this early phase of development is that when positivity thresholds were defined, they were usually established *a posteriori*, without confirmatory follow-up studies.

The scarce clinical utility for making a precise diagnosis of NASH in the absence of a clear practical consequence (e.g. access to treatment) most likely limited the research on non-invasive tests at an academic level. On the other hand, recent guidelines recommend having a histological diagnosis of NASH.[2,9] Indeed, resolution of NASH is presently considered a major endpoint in clinical trials, which will hopefully soon lead to the approval of the first NASH therapies,[9] providing a strong rationale for the non-invasive assessment of this condition.

This new perspective demands that research on non-invasive tests for diagnosis of NASH enter a new phase, starting from those tests which have emerged as promising thanks to their initial accuracy, are based on feasible techniques and have no or minimal direct harms of testing.

Even if a feasibility analysis of the different techniques is beyond the scope of this review, some issues are self-evident: work load and costs are higher for MR than for US, and techniques which require contrast media administration or complex post-processing, for example MRS, have additional costs. As for direct harms, they may include radiation dose (CT, scintigraphy) and contrast media administration (gadoxetic acid and SPIO/USPIO). Other techniques are substantially free of direct harms.

Based on accuracy data, the most promising tests among techniques which are relatively feasible and harmless are US and MR imaging, including both elastography (shear wave-

based elastography, MRE) and non-elastographic techniques (some US scores, multiparametric MRI, susceptibility-weighted imaging), which can possibly be combined. Their combination with circulating biomarkers may also provide an added value in terms of accuracy, and research is also very active in this field.[47] A clinically applicable diagnostic algorithm will probably comprise scoring system and circulating biomarkers to be used to select high-risk patients who could benefit from a combination of imaging tests.[48]

Some of these techniques may have intrinsic limitations for NASH diagnosis. Elastographic techniques have been validated to assess fibrosis. Even if liver stiffness increase may also be due to inflammation, there is the possibility that these techniques have an acceptable accuracy in diagnosing NASH as a consequence of the strong association between the presence of NASH and fibrosis. Hence, they could have intrinsic limit in sensitivity, not identifying NASH without fibrosis. However, Costa-Silva et al. observed a similar accuracy of MRE for NASH diagnosis in patients with and without fibrosis.[29]

Techniques aimed at quantifying fat accumulation have failed to reach a mature stage of validation in NASH diagnosis. Steatosis is a necessary condition for both NASH and NAFLD, but assessment of hepatic fat amount may not be sufficient to identify patients with inflammation. Similarly, US scores mostly evaluating liver hyperechogenicity [22] present high referral rates and low positive predictive values to obtain high sensitivity. Preperitoneal fat area and perihepatic adipose tissue thickness, evaluated by means of MRI and US,[33,42] likewise showed high sensitivities and relatively low specificities at the proposed thresholds. Indeed, these are not direct measures of inflammation but rather indicators of visceral adiposity.

Some limitations of this review must be acknowledged. First, the search algorithm included only some techniques specifically reported in the string. Second, the choice not to pool data from the few studies that analyzed the same technique, but with different procedures, thresholds and populations, was somewhat arbitrary.

In conclusion, several imaging techniques have been tested for accuracy in NASH diagnosis. US and MR imaging, including both elastography and non-elastographic techniques, have shown promising accuracy and have no direct harms. Their combination with circulating biomarkers may provide efficient algorithms, thereby contributing to increasing diagnostic accuracy. However, the studies were conducted in limited series of patients, with different clinical features and selection criteria, using various NASH definitions and lacking independent validation. The picture of this early stage of development underlines the need for large collaborative multicenter studies with prospective design and clear definitions of outcomes, which would allow a direct comparison of the most promising imaging and biomarker approaches for NASH diagnosis.

ACKNOWLEDGEMENTS

We thank Jacqueline Costa for the English language editing.

REFERENCES

1. Estes C, Razavi H, Loomba R, Younossi Z, Sanyal AJ. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. Hepatology. 2018;67(1):123–33. doi:10.1002/hep.29466
2. Chalasani N, Younossi Z, Lavine JE, et al. The Diagnosis and Management of Nonalcoholic Fatty Liver Disease : Practice Guidance From the American Association for the Study of Liver Diseases. Hepatology. 2018;67(1):328–57. doi: 10.1002/hep.29367
3. Sumida Y, Nakajima A, Itoh Y. Limitations of liver biopsy and non-invasive diagnostic tests for the diagnosis of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. World J Gastroenterol. 2014;20(2):475–85. doi: 10.3748/wjg.v20.i2.475
4. Issa D, Patel V, Sanyal A. Future therapy for non-alcoholic fatty liver disease. Liver Int. 2018;38(Suppl 1):56–63. doi: 10.1111/liv.13676
5. Park C, Nguyen P, Hernandez C, et al. Magnetic Resonance Elastography vs Transient Elastography in Detection of Fibrosis and Noninvasive Measurement of Steatosis in Patients with Biopsy-proven Nonalcoholic Fatty Liver Disease. Gastroenterology. 2017;152(3):598–607. doi: 10.1053/j.gastro.2016.10.026
6. Dulai P, Singh S, Patel J, et al. Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: Systematic review and meta-analysis. Hepatology. 2017;65(5):1557–65. doi: 10.1002/hep.29085
7. Crossan C, Tsochatzis E, Longworth L, et al. Cost-effectiveness of non-invasive methods for assessment and monitoring of liver fibrosis and cirr hosis in patients with chronic liver disease: systematic review and economic evaluation. Heal Technol Assess. 2015;19(9):1–409. doi: 10.3310/hta19090

8. Xiao G, Zhu S, Xiao X, Yan L, Yang J, Wu G. Comparison of laboratory tests, ultrasound, or magnetic resonance elastography to detect fibrosis in patients with nonalcoholic fatty liver disease: A meta-analysis. Hepatology. 2017;66(5):1486–501. doi: 10.1002/hep.29302

9. European Association for the Study of the Liver (EASL); European Association for the Study of Diabetes (EASD); European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease.. J Hepatol. 2016;64(6):1388–402. doi: 10.1016/j.jhep.2015.11.004

10. Bedossa P, Patel K. Biopsy and Noninvasive Methods to Assess Progression of Nonalcoholic Fatty Liver Disease. Gastroenterology. 2016;150(8):1881–1822. doi: 10.1053/j.gastro.2016.03.008.

11. Whiting P, Rutjes A, Westwood M,et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36. doi: 10.7326/0003-4819-155-8-201110180-00009

12. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol. 2011 Apr;64(4):380-2. doi: 10.1016/j.jclinepi.2010.09.011.

13. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLOS Med. 2009;6(7):e1000097. doi: 10.1371/journal.pmed.1000097

14. Matteoni CA, Younossi ZM, Gramlich T, Boparai N, Liu YC, McCullough AJ. Nonalcoholic fatty liver disease: a spectrum of clinical and pathological severity. Gastroenterology. 1999;116:1413-1419.

15. Brunt EM1, Janney CG, Di Bisceglie AM, Neuschwander-Tetri BA, Bacon BR. Nonalcoholic Steatohepatitis : A Proposal for Grading and Staging the Histological

Lesions. Am J Gastroenterol. 1999;94(9):2467–74.

16. Bedossa P, Poitou C, Veyrie N, et al. Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. Hepatology. 2012;56:1751-1759. doi: 10.1002/hep.25889

17. Kleiner DE, Brunt EM, Van Natta M, et al. Design and Validation of a Histological Scoring System for Nonalcoholic Fatty Liver Disease. Hepatology. 2005;41(6):1313–21.

18. Imajo K, Kessoku T, Honda Y, et al. Magnetic Resonance Imaging More Accurately Classifies Steatosis and Fibrosis in Patients With Nonalcoholic Fatty Liver Disease Than Transient Elastography. Gastroenterology. 2016;150:626–37. doi: 10.1053/j.gastro.2015.11.048

19. Guzman-Aroca F, Frutos-Bernal M, Bas A, et al. Detection of non-alcoholic steatohepatitis in patients with morbid obesity before bariatric surgery: preliminary evaluation with acoustic radiation force impulse imaging. Eur Radiol. 2012;22(11):2525–32. doi: 10.1007/s00330-012-2505-3

20. Liang RJ, Wang HH, Lee WJ, Liew PL, Lin JT, Wu MS. Diagnostic Value of Ultrasonographic Examination for Nonalcoholic Steatohepatitis in Morbidly Obese Patients Undergoing Laparoscopic Bariatric Surgery. Obes Surg. 2007;17:45–56.

21. Abrigo JM, Shen J, Wong VW, et al. Non-alcoholic fatty liver disease : Spectral patterns observed from an in vivo phosphorus magnetic resonance spectroscopy study. J Hepatol. 2014;60:809–15. doi: 10.1016/j.jhep.2013.11.018

22. Petrick A, Benotti P, Wood G, et al. Utility of Ultrasound, Transaminases, and Visual Inspection to Assess Nonalcoholic Fatty Liver Disease in Bariatric Surgery Patients. Obes Surg. 2015;25(12):2368–75. doi: 10.1007/s11695-015-1707-6

23. Chen J, Talwalkar JA, Yin M, Glaser KJ, Sanderson SO, Ehman RL.. Early detection

of nonalcoholic steatohepatitis in patients with nonalcoholic fatty liver disease by using MR elastography. Radiology. 2011;259(3):749–56. doi: 10.1148/radiol.11101942

24. Gallego-Durán R, Cerro-Salido P, Gomez-Gonzalez E, et al. Imaging biomarkers for steatohepatitis and fibrosis detection in non-alcoholic fatty liver disease. Sci Rep. 2016;6:31421. doi: 10.1038/srep31421

25. Naganawa S, Enooku K, Tateishi R, et al. Imaging prediction of nonalcoholic steatohepatitis using computed tomography texture analysis. Eur Radiol. 2018;28(7):3050–8. doi: 10.1007/s00330-017-5270-5

26. Eddowes P, McDonald N, Davies N, et al. Utility and cost evaluation of multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease. Aliment Pharmacol Ther. 2018;47:631–44. doi: 10.1111/apt.14469

27. Lee HW, Park SY, Kim SU, et al. Discrimination of Nonalcoholic Steatohepatitis Using Transient Elastography in Patients with Nonalcoholic Fatty Liver Disease. PLoS One. 2016;11(6):e0157358. doi: 10.1371/journal.pone.0157358

28. Fierbinteanu Braticevici C, Sporea I, Panaitescu E, Tribus L. Value Of Acoustic Radiation Force Impulse Imaging Elastography For Non-Invasive Evaluation Of Patients With Nonalcoholic Fatty Liver Disease. Ultrasound Med Biol. 2013;39(11):1942–50. doi: 10.1016/j.ultrasmedbio.2013.04.019

29. Costa-Silva L, Ferolla SM, Lima AS, Vidigal PVT, Ferrari TCA. MR elastography is effective for the non-invasive evaluation of fibrosis and necroinflammatory activity in patients with nonalcoholic fatty liver disease. Eur J Radiol. 2018;98:82–9. doi: 10.1016/j.ejrad.2017.11.003

30. Loomba R, Wolfson T, Ang B, et al. Magnetic Resonance Elastography Predicts Advanced Fibrosis in Patients With Nonalcoholic Fatty Liver Disease: A Prospective

Study. Hepatology. 2014;60:1920–1928. doi: 10.1002/hep.27362

31. Loomba R, Cui J, Wolfson T, et al. Novel 3D magnetic resonance elastography for the noninvasive diagnosis of advanced fibrosis in NAFLD: A prospective study. Am J Gastroenterol. 2016;111(7):986–94. doi: 10.1038/ajg.2016.65
32. Ballestri S, Lonardo A, Romagnoli D, et al. Ultrasonographic fatty liver indicator , a novel score which rules out NASH and is correlated with metabolic parameters in NAFLD. Liver Int. 2012;32(8):1242–52. doi: 10.1111/j.1478-3231.2012.02804.x
33. Lirussi F, Vitturi N, Azzalini L, et al. Perihepatic Adipose Tissue Thickness: a New Non-Invasive Marker of NAFLD? J Gastrointestin Liver Dis. 2009;18(1):61–6.
34. Tarantino G, Conca P, Pasanisi F, et al. Could inflammatory markers help diagnose nonalcoholic steatohepatitis? Eur J Gastroenterol Hepatol. 2009;21(5):504–11. doi: 10.1097/MEG.0b013e3283229b40
35. Zardi EM, De Sio I, Ghittoni G, et al. Which Clinical and Sonographic Parameters May Be Useful to Discriminate NASH from Steatosis? J Clin Gastroenterol. 2011;45:59–63. doi: 10.1097/MCG.0b013e3181dc25e3.
36. Iijima H, Moriyasu F, Tsuchiya K, et al. Decrease in accumulation of ultrasound contrast microbubbles in non-alcoholic steatohepatitis. Hepatol Res. 2007;37:722–30.
37. Kim T, Jun HY, Kim K, et al. Hepatic Alanine Differentiates Nonalcoholic Steatohepatitis From Simple Steatosis in Humans and Mice: A Proton MR Spectroscopy Study With Long Echo Time. J Magn Reson Imaging. 2017;46:1298–310. doi: 10.1002/jmri.25673.
38. Pavlides M, Banerjee R, Tunnicliffe EM, et al. Multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease severity. Liver Int. 2017;37:1065–73. doi: 10.1111/liv.13284
39. Leporq B, Lambert SA, Ronot M, Vilgrain V, Van Beers BE. Simultaneous MR

quantification of hepatic fat content, fatty acid composition, transverse relaxation time and magnetic susceptibility for the diagnosis of non-alcoholic steatohepatitis. NMR Biomed. 2017; 30(10). doi: 10.1002/nbm.3766

40. Parente DB, Paiva FF, Neto JA, 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. 2015;10(5):e0125653. doi: 10.1371/journal.pone.0125653

41. Dillman JR, Trout AT, Costello EN, et al. Quantitative Liver MRI-Biopsy Correlation in Pediatric and Young Adult with Nonalchohlic Fatty Liver Disease: Can One Be Used to Predict the Other? AJR. 2018;210:166–74. doi: 10.2214/AJR.17.18446

42. Parente D, Oliveira Neto J, Brasil P, et al. Preperitoneal fat as a non-invasive marker of increased risk of severe non-alcoholic fatty liver disease in patients with type 2 diabetes. J Gastroenterol Hepatol. 2018;33:511–7. doi: 10.1111/jgh.13903

43. Bastati N, Feier D, Wibmer A, et al. Noninvasive Differentiation of Simple Steatosis and Steatohepatitis by Using Gadoxetic Acid–enhanced MR Imaging in Patients with Disease: A Proof-of-Concept Study. Radiology. 2014;271(3):739–47. doi: 10.1148/radiol.14131890

44. Smits L, Coolen B, Panno M, 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. 2016;278(3):782–91. doi: 10.1148/radiol.2015150952

45. Tomita K, Tanimoto A, Irie R, et al. Evaluating the Severity of Nonalcoholic Steatohepatitis With Superparamagnetic Iron Oxide-Enhanced Magnetic Resonance Imaging. J Magn Reson Imaging. 2008;28:1444–50. doi: 10.1002/jmri.21596

46. Kikuchi M, Tomita K, Nakahara T, et al. Utility of quantitative Tc-phytate

scintigraphy to diagnose early-stage non-alcoholic steatohepatitis. Scand J Gastroenterol. 2009;44:229–36. doi: 10.1080/00365520802433249.

47. Wong VW, Adams LA, de Lédinghen V, Wong GL, Sookoian S. Noninvasive biomarkers in NAFLD and NASH - current progress and future promise. Nat Rev Gastroenterol Hepatol. 2018;15(8):461–78. doi: 10.1038/s41575-018-0014-9.

48. Yoneda M, Imajo K, Nakajima A. Non-Invasive Diagnosis of Nonalcoholic Fatty Liver Disease. Am J Gastroenterol. 2018 May 1:1409-1411. doi: 10.1038/s41395-018-0170-0.

TABLES

Elastographic techniques

Study	Study design and index test	Population and NASH prevalence	NASH definition	Accuracy simple steatosis vs NASH
Transient Elastography (TE) – liver stiffness (LS) Coefficient attenuation parameter (CAP)				
Eddowes 2018[26]	Prospective LS	50 patients; 38(76%) with NASH, 47 with reliable TE	Steatosis, lobular inflammation and ballooning	AUROC=0.82(0.70-0.94) AUROC for NAS≥5=0.74(0.59-0.89)
Imajo 2016[18]	Prospective LS; CAP	142 patients; 108(76%) with NASH, 127 with reliable TE	Steatosis, inflammation, ballooning, and pericellular/perisinusoidal fibrosis	AUROC=0.80(0.73–0.88) † AUROC for NAS≥5=0.65(0.54-0.77) †
Lee 2016[27]	Retrospective LS; CAP	183 patients 94(51.4%) with NASH	Steatosis, inflammation and ballooning; NAS≥5	LS>7 kPa: AUROC=0.751(0.677–0.824); sensitivity=86.2%, specificity=58.4% CAP>250 dB/m: AUROC=0.743(0.669–0.816), sensitivity=96%, specificity=49% Score based on LS, CAP and ALT: AUROC=0.812(0.724–0.880)
Park 2017[5]	Prospective LS	104 patients 76(76%) with NASH ‡	NAS≥2	AUROC=0.35(0.22-0.49) Cut-off>5.6 KPa: sensitivity=61.1%, specificity=59.1%, PPV=83%, NPV=31.7%
Acoustic Radiation Force Impulse (ARFI) – shear wave velocity (SWV)				
Fierbinteanu Braticevici 2013[28]	Prospective SWV	64 patients 43(67%) with NASH	Brunt 1999/Kleiner 2005 criteria. Patients divided into simple steatosis and NASH, *borderline patients excluded*.	AUROC=0.87 Cut-off>1.10 m/s: sensitivity=77%, specificity=72%, PPV=85%, NPV=60%

Guzman-Aroca 2012[19]	Prospective SWV	32 bariatric patients 24(75%) with NASH/fibrosis (18 with inflammation and 6 with fibrosis)	Matteoni 1998 criteria. Patients categorized as simple steatosis, inflammation and fibrosis. *Comparisons between SS and NASH/fibrosis.*	NASH and/or fibrosis vs simple steatosis: AUROC=0.9 Cut-off 1.3 m/s: sensitivity=85%, specificity=83%, PPV=89%, NPV=77%
Magnetic Resonance Elastography (MRE) – liver stiffness (LS)				
Chen 2011[23] (2D MRE)	Retrospective LS	58 patients 36(72%) with NASH/fibrosis (7 inflammation and 29 fibrosis)	Brunt 1999. Patients categorized as simple steatosis, inflammation without fibrosis, and NAFLD with fibrosis, the latter two classified as NASH.	AUROC=0.93 Cut-off>2.74 KPa: sensitivity=94%, specificity=73%, PPV=85%, NPV=89% Cut-off>2.90 KPa: sensitivity=83%, specificity=82%, PPV=88%, NPV=75%
Costa-Silva 2018[29] (2D MRE)	Prospective LS	49 patients 25(51%) with NASH	NAS≥5	AUROC=0.79 Cut-off 3.24 Kpa: sensitivity=72%, specificity=88%, PPV=86%, NPV=72%. in fibrosis=0 patients (n=21): AUROC=0.78 Cut-off 3.22 kPa: sensitivity=69%, specificity=87%
Imajo 2016[18] (2D MRE)	Prospective LS	142 patients; 108(76%) with NASH	Steatosis, inflammation, ballooning and pericellular/perisinusoidal fibrosis	AUROC=0.81 § AUROC for NAS≥5=0.77 §
Loomba 2014[30] (2D MRE)	Prospective LS	117 patients 106(91%) with NASH	Kleiner 2005. Borderline with definite NASH.	AUROC=0.73 Cut-off 3.26 Kpa: sensitivity=42%; specificity=92%; PPV=95%;

				NPV=32%
Loomba 2016[31] (2D and 3D MRE)	Prospective LS	100 patients 87(87%) with NASH	Kleiner 2005. Borderline with definite NASH.	2D MRE (60 Hz): AUROC=0.75; optimal cut-off=2.92 Kpa; 3D MRE (60 Hz): AUROC=0.76; optimal cut-off=2.42 Kpa; 3D MRE (40 Hz): AUROC=0.74; optimal cut-off=1.93 KPa
Park 2017[5] (2D MRE)	Prospective LS	104 patients 76(76%) with NASH ‡	NAS≥2	AUROC=0.70 Cut-off>2.53 KPa: sensitivity=63.9%, specificity=68.2%, PPV=86.8%, NPV=36.6%

Table 1: Summary of included studies with one or more elastographic techniques as index test and diagnostic accuracy as outcome. AUROC: area under the receiver operating characteristic curve. NAS, NAFLD Activity Score. † For the combination of liver stiffness and CAP; ‡ Histological data reported for 100/104 patients; § For the combination of MRE and Proton Density Fat Fraction.

US non-elastographic techniques

Study	Study design and index test	Population and NASH prevalence	NASH definition	Accuracy simple steatosis vs NASH
US B-mode parameters and scores				
Ballestri 2012[32]	Prospective US-fatty liver indicator (US-FLI) (2-8): liver/kidney contrast (2–3), US posterior attenuation (0–1), vessel blurring (0–1), difficult visualization of gallbladder wall (0–1) or diaphragm (0–1), focal sparing (0–1)	53 patients; 35(66%) with NASH	Steatosis, lobular inflammation and ballooning; severe NASH for NAS≥ 5	AUROC=0.76 for NASH; 0.80 for severe NASH. US-FLI<4 ruled out severe NASH with NPV=94%; specificity=46%.
Liang 2007[20]	Prospective US fatty score (FS) (0–8): parenchymal echogenicity, far gain attenuation, gallbladder wall blurring, portal vein wall blurring, and	101 obese bariatric patients; 72(71%) with NASH	Fibrosis (≥grade 1) or acinar zone 3 hepatocellular injury with ballooning (≥grade 2)	FS: AUC=0.79; cut-off 7; sensitivity=81%; specificity=66%; accuracy=76%; PPV=85%; NPV=58% MFS: AUC=0.82; cut-off 3; sensitivity=72%; specificity=86%; accuracy=76%; PPV=93%; NPV=56%

	hepatic vein blurring. Modified FS (MFS) (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			
Lirussi 2009[33]	Prospective US PATT (perihepatic adipose tissue thickness)	65 patients (33 with liver biopsy); 27(82%) with NASH	Brunt 1999. Borderline with definite NASH	Cut-off 11.8 mm: sensitivity=100%, Specificity=50%, AUROC=75%. To predict necro-inflammatory activity grading: sensitivity=80%, specificity=50%, AUROC=60%
Petrick 2015[22]	Prospective US-Fatty liver (mild, moderate, or severe according to the fall in echo amplitude, extent of liver/kidney discrepancy and	513 bariatric patients 146(28%) with steatohepatitis; 164(32%) with NASH.	Brunt 1999. Steatohepatitis defined as lobular inflammation; NASH defined as steatohepatitis, fibrosis or	For steatohepatitis: US fatty liver (mild+): sensitivity=89%; specificity=45%; PPV=39%; NPV=91%; Accuracy=58%

	of echo loss from portal vein)		cirrhosis	
Tarantino 2009[34]	Prospective Spleen longitudinal diameter	83 patients; 43(52%) with NASH	Kleiner 2005. Lobular inflammation 0-3, no further specified NASH definition	AUROC=0.920 Cut-off 116 mm: sensitivity=88%, specificity=95%
Zardi 2011[35]	Retrospective US score (0-6): echo amplitude attenuation (0–2), focal fat sparing (0–1), splenic diameter (0–3).	94 patients; 74(79%) with NASH	Steatosis, lobular inflammation and ballooning.	Cut-off≥5: sensitivity=74%, specificity=66%; only echo attenuation and focal fat sparing (cut-off=1): sensitivity=92%, specificity=75%.
Contrast-Enhanced Ultrasound (CEUS)				
Iijima 2007[36]	Prospective CEUS Signal intensities 5 and 20 minutes after Levovist administration	66 patients (liver biopsy in 31 patients: 21 with NASH; *in the remaining 35 NASH was clinically excluded*) + 10 healthy volunteers †	Brunt 1999. NASH for presence of parenchymatitis independently of fibrosis	Signal intensity 5 minutes Cut-off=137.8: sensitivity=100%, specificity=95%, accuracy=80%. Signal intensity 20 minutes Cut-off=43.6: sensitivity, specificity and accuracy=100%.

Table 2: Summary of included studies with one or more US non-elastography techniques as index test and diagnostic accuracy as outcome. AUROC: area under the receiving operating characteristic curve. NAS: NAFLD Activity Score. † not clear whether included in analysis.

MR non-elastographic techniques

Study	Study design and index test	Population and NASH prevalence	Definition of NASH	Accuracy simple steatosis vs NASH
^{1}H-MRS and/or ^{31}P-MRS metabolites				
Abrigo 2014[21] (*^{31}P-MRS)*	Prospective Nucleotide Triphosphate (α peak)/Triphosphate (αNTP/TP)	132 patients 95(72%) with NASH	Matteoni 1998. NASH for type 3 and 4 (fat accumulation and ballooning ± Mallory hyaline or fibrosis)	α-NTP/TP: AUROC=0.71 Cut-off≤10.57%: sensitivity=28%; specificity=91%; PPV=78%; NPV=43% Cut-off≤16.36%: sensitivity=91%; specificity=16%; PPV=65%; NPV=50%.
Kim 2017[37] (long echo time ^{1}H-MRS)	Prospective Alanine (Ala), lactate+trygliceride (Lac+TG)	26 patients; 11(42%) with NASH	NAS≥5	Ala: AUROC=1.00 Cut-off>16.04%: sensitivity=100%, specificity=100% Lac+TG: AUROC=0.78 Cut-off>360.8%: sensitivity=82%, specificity=67%
Multiparametric MRI (Liver MultiScan)- corrected T1 (cT1), Liver Inflammation and Fibrosis (LIF) score				
Eddowes 2018[26]	Prospective T1 corrected for T2* (cT1)	50 patients 38(76%) with NASH	Lobular inflammation and ballooning	AUROC for NASH vs SS=0.69 AUROC for NAS≥5 vs <5=0.74
Pavlides 2017[38]	Prospective LIF score (0-4) based on cT1 cut-offs.	71 patients 46(65%) with NASH	Steatosis, ballooning, lobular inflammation	AUROC=0.80 Cut-off 1.4: sensitivity=91%, specificity=52%
Diffusion weighted (DW) MRI and Intravoxel Incoherent Motion DW MRI (D, D, f)*				
Parente 2015[40]	Prospective Pure molecular-based (D), perfusion-related (D*), and vascular (f) Fractions	59 T2DM patients; 22(37%) with NASH	Steatosis, lobular inflammation and ballooning	-D: AUROC=0.742; cut-off 0.760: sensitivity=69% specificity=66%; -D*: AUROC=0.678; cut-off 41.45: sensitivity=68% specificity=71%; -f: AUROC=0.607; cut-

				off 34.23: sensitivity=49% specificity=70%.
Quantitative susceptibility MRI				
Leporq 2017[39]	Retrospective Susceptibility (ppm)	32 patients; 20(62.5%) with NASH	Steatosis, ballooning, lobular inflammation	AUROC=0.91
MRI optical analysis				
Gallego-Duran 2016[24]	Prospective NASHMRI score obtained from most predicting estimators	126 patients (estimation cohort n=39 and validation cohort n=87); 65(51%) with NASH	Kleiner 2005. Ballooning and inflammation.	NASHMRI score: -estimation cohort: AUROC=0.88. Best cut-off>0.5: sensitivity=87%, specificity=74%, PPV=80%, NPV=82% -validation cohort: AUROC=0.83. Cut-off>0.5: sensitivity=87%, specificity=60%, PPV=71% and NPV=81%.
Gadoxetic acid-enhanced MRI				
Bastati 2014[43]	Retrospective Relative Enhancement in hepatobiliary phase	81 patients; 35(43%) with NASH	NASH for activity≥2 and steatosis≥1 with any fibrosis	AUROC=0.85 Cut-off≤1.24: sensitivity=97%; specificity=63%
SPIO/USPIO-enhanced MRI				
Smits 2015[44] (USPIO)	Prospective Difference (Δ) in R2* between contrast-enhanced and baseline	24 patients (6 simple steatosis patients not biopsy-proven) 13(54%) with NASH	NAS≥5 when steatosis, inflammation and ballooning present	AUROC=0.87 Cut-off<45.5 sec^{-1}: sensitivity=77%; specificity=91%. Cut-off<58.3 sec^{-1}: sensitivity=85%; specificity=73%.
Tomita 2008[45] (SPIO)	Prospective Relative decrease in T2 (%T2) and time constant (T)	19 patients; 10(53%) with NASH	NAS≥5	T: AUROC=0.79 Cut-off=42.8: specificity=67%, sensitivity=100%, PPV=77%, NPV=100%.

				%T2: AUROC=0.83 Cut-off=32.5: specificity=73%, sensitivity=88%, PPV=70%, NPV=89%.
MRI Liver Volume				
Dillman 2018[41]	Retrospective Liver Volume	69 children and young adults ≤21 years old; 37(54%) with NASH	NAS≥5	AUC=0.741
MRI preperitoneal fat area				
Parente 2018[42]	Prospective Preperitoneal fat area (cm^2)	66 T2DM patients; 23(35%) with NASH	Steatosis, ballooning and lobular Inflammation	Cut-off=5: sensitivity=93%; specificity=55%

Table 3: Summary of included studies with one or more MR non-elastographic techniques as index test and diagnostic accuracy as outcome. AUROC: area under the receiving operating characteristic curve. T2DM: Type 2 Diabetes Mellitus; NAS: NAFLD Activity Score.

Other techniques

Study	Study design and index test definition	Population and NASH prevalence	Definition of NASH	Accuracy simple steatosis vs NASH
Computed Tomography (CT)				
Naganawa 2018[25]	Retrospective Non-Contrast-Enhanced CT texture features; logistic models for NASH from the most predictive features	88 patients (learning dataset=53 patients and validation dataset=35 patients). Prevalence of NASH not reported.	NAS≥3	Patients without high suspicion of fibrosis: NASH model based on mean0 and skewness2, with cut-off=0.45: AUROC=0.93 and 0.94 in learning and validation datasets; accuracy=94%, specificity=92%, sensitivity=100%, PPV=100%, NPV=80%. Patients with high suspicion of fibrosis: NASH model based on mean0 and kurtosis4, with cut-off=0.81: AUROC=0.81 and 0.60 in learning and validation datasets, accuracy=42%, specificity=31%, sensitivity=100%, PPV=100%, NPV=21%.
Liver Scintigraphy				
Kikuchi 2009[46] (Tc99m-phytate colloid scintigraphy)	Prospective Liver-to-spleen uptake ratio	37 patients; 29(78%) with definite NASH.	Kleiner 2005. Definite NASH for NAS≥5 (no patient with borderline NASH)	AUC=0.82 Cut-off value=2.93: specificity=75%, sensitivity=100%, PPV=94%, NPV=100%

Table 4: Summary of included studies with techniques other than elastography, US and MR as index test and diagnostic accuracy as outcome. AUROC: area under the receiving operating characteristic curve. NAS: NAFLD Activity Score.

	Potential harms		Resource consumption			Operator-dependence	Accuracy			Stage of development	
	Contrast media or Radiotracer	Radiation dose	Human	Technology: hardware	Technology: software		Certainty of the evidence[†]	Sensitivity[‡]	Specificity[‡]	Harmonized procedures[§]	Defined and agreed positivity threshold[§]
Transient Elastography	no	no	low	low	low	yes	very low to low	varies	low	yes	no
US shear wave-based elastography	no	no	low	low	low	yes	low	fair	fair	yes	no/yes
MR Elastography	no	no	high	very high	high	no	very low	varies	fair	yes	no/yes
US non-elastographic scores and parameters	no	no	low	low	low	yes	very low	varies	varies	no	no
Contrast-enhanced US	yes	no	low	low	low	yes	very low	good	good	no	no
MR Spectroscopy	no	no	very high	high	high	no	very low	varies	varies	no	no
Multiparametric-MRI	no	no	high	high	high	no	low to very low	good	low	yes	no
IVIM-DW-MRI	no	no	high	high	high	no	very low	varies	fair	yes	no
Susceptibility-weighted MRI	no	no	high	high	high	no	very low	good	fair	no	no

MRI optical analysis	no	no	very high	high	very high	no	low to moderate	fair	fair	no	no
MRI morphological parameters	no	no	high	high	high	no	very low	fair	varies	no	no
Contrast-enhanced MR	yes	no	high	high	high	no	very low	fair	fair	no	no
CT texture analysis	no	yes	very high	high	very high	no	very low	good	varies	no	no
Scintigraphy	yes	yes	high	high	high	no	very low	good	fair	no	no

Table 5: Advantages and disadvantages of the techniques under evaluation in terms of potential harms, resource consumption, operator-dependence, accuracy, and stage of development. † level of evidence was classified according to GRADE criteria. ‡ According to table 1 to 4, low was assigned if the results with different procedures were substantially <=60%, fair if >60% to 90%, good >90%; otherwise we reported varies. § Procedures were considered harmonized and positivity thresholds defined and agreed on when more than one study reported on the same techniques with similar procedures and positivity thresholds (no/yes was reported when cut-off values were similar among studies, even if data pooling was not possible due to other sources of heterogeneity).

FIGURE LEGENDS

Fig. 1: PRISMA 2009 Flow Diagram of searched, screened and included studies.

Fig. 2: Diagram depicting all evaluated imaging techniques, subdivided into 4 categories (elastography, ultrasonography, magnetic resonance, other). Studies assessing a single technique are reported outside the circle, while studies reported inside the circle compared two or more techniques, linked to each study through lines. TE, transient elastography; ARFI, acoustic radiation force impulse; MRE, magnetic resonance elastography; CEUS, contrast-enhanced ultrasound; MRS, magnetic resonance spectroscopy; IVIM, intravoxel incoherent motion; DWI, diffusion-weighted imaging. Studies without accuracy results are cited in Supplementary References.

Fig. 3: Classification of index tests based on the kind of feature studied (physical properties - liver stiffness, anatomical features, tissue composition or functional features). For each index test a brief explanation is reported together with the studies addressing each modality.