BMC Gastroenterology

https://doi.org/10.1186/s12876-026-05236-z

Article in Press

Diagnostic performance of quantitative ultrasound techniques for the diagnosis and grading of hepatic steatosis in metabolic dysfunction-associated steatotic liver disease: a systematic review and network meta-analysis

Received: 25 May 2026

Accepted: 14 August 2026

Published online: 24 August 2026

Cite this article as: Liu X., Liu F., Zhai X. *et al.* Diagnostic performance of quantitative ultrasound techniques for the diagnosis and grading of hepatic steatosis in metabolic dysfunction-associated steatotic liver disease: a systematic review and network meta-analysis. *BMC Gastroenterol* (2026). https://doi.org/10.1186/s12876-026-05236-z

Xingya Liu, Funing Liu, Xuechao Zhai, Wenrui Liu & Hongbo Wang

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

Diagnostic Performance of Quantitative Ultrasound Techniques for the Diagnosis and Grading of Hepatic Steatosis in Metabolic Dysfunction-Associated Steatotic Liver Disease: A Systematic Review and Network Meta-Analysis

Xingya Liu[1], Funing Liu[1], Xuechao Zhai[2], Wenrui Liu[3], Hongbo Wang[4,5*]**

[1] School of First Clinical Medicine, Shenyang Medical College, Shenyang, Liaoning 110034, China

[2] School of Mathematics and Information Science, Hebei University, Baoding, Hebei 071051, China

[3] Central Hospital of Shenyang Sujiatun, Shenyang, Liaoning, 110034, China

[4] Department of Human Anatomy, School of Basic Medicine, Shenyang Medical College, Shenyang, Liaoning, 110034, China

[5]Liaoning Province Key Laboratory for Phenomics of Human Ethnic Specificity and Critical Illness, Shenyang, Liaoning, 110034, China

[*]**Corresponding author:** Hongbo Wang, No.146, Huanghe North Street, Shenyang, Liaoning, 110034, China.

Email: wanghongbo@symc.edu.cn

Tel: 18640249580

[**]**Co-Corresponding author:** Wenrui Liu, Email: 86262463@qq.com

ARTICLE IN PRESS

Abstract

Objective: Quantitative ultrasound-derived techniques are core non-invasive tools assessing liver steatosis in metabolic dysfunction-associated steatotic liver disease (MASLD), providing direct fat fraction measurements for screening, grading, and monitoring. This study compares the diagnostic performance of varied quantitative ultrasound-derived approaches relative to magnetic

ARTICLE IN PRESS

resonance imaging-derived proton density fat fraction (MRI-PDFF) via a network meta-analysis (NMA).

Methods: PubMed, Embase, the Cochrane Library, and Web of Science were systematically searched until December 7, 2025, for English diagnostic accuracy studies comparing quantitative ultrasound-based techniques with MRI-PDFF for MASLD grading. The primary outcomes extracted were sensitivity (SEN) and specificity (SPE). Study quality was assessed via the QUADAS-2 tool. The primary analysis employed a Bayesian arm-based bivariate random-effects network meta-analysis of diagnostic accuracy implemented in RStan, jointly estimating the pooled SEN, pooled SPE, diagnostic odds ratio (DOR), and superiority index (SI) of each technique within the same model. In addition, exploratory univariate network meta-analyses were performed for SEN and SPE separately using gemtc to calculate the relative sensitivity, relative specificity, and corresponding SUCRA values.

Results: 17 studies from 10 countries were included, comprising 1,904 MASLD patients and 1,041 controls and covering 17 quantitative ultrasound-derived techniques. The overall risk of bias (RoB) was moderately low. Except for the SEN analysis of S1 steatosis, no significant publication bias was detected across the remaining grading analyses, although evidence of small-study effects was observed. The bivariate random-effects model demonstrated that SWA (0.96), UDFF (0.92), and TAI (0.98) achieved the highest SEN for diagnosing S1, S2, and S3 steatosis, respectively (e.g., S1 (mild):⩾S1 vs S0, S2 (moderate):⩾S2 vs S0-S1, S3 (severe):⩾S3 vs S0-S2). For SPE, RFM + ACE + HI showed superior performance for S1 (0.95) and S2 (0.91), whereas UDFF/Deep-USFF demonstrated the highest SPE for S3 steatosis (0.89). The superiority index, which comprehensively evaluates overall diagnostic performance, showed that RFM+ACE+HI performed remarkably well in S1 (SI=14.25) and S2 (SI=15.72), whereas

ARTICLE IN PRESS

Deep-USFF showed excellent performance in S3 (SI=14.61). The diagnostic odds ratios indicated that SWA exhibited the highest DOR in S1 (373,719.55, 95% CrI: 8.08 – 853,672.04) and S2 (234.01, 95% CrI: 1.86 – 1,077.23), while TAI showed the highest DOR in S3 (2,498,255.36, 95% CrI: 34.43 – 1,309,055.51). Some estimates were accompanied by extremely wide credible intervals (owing to zero-event studies), indicating uncertainty and warranting cautious interpretation.

Conclusion: This NMA indicates that quantitative ultrasound-derived techniques exhibit distinct diagnostic strengths for grading hepatic steatosis in MASLD. SWA, UDFF, and TAI demonstrated superior SEN for detecting mild (S1), moderate (S2), and severe (S3) steatosis, respectively, whereas RFM + ACE + HI and UDFF/Deep-USFF showed greater SPE for their corresponding steatosis grades. This study relied primarily on the composite indicators from the bivariate model as the core basis for evaluation, with some indicators exhibiting certain uncertainties due to the presence of zero-event studies. The SUCRA values calculated based on SEN and SPE separately reflect only the relative ranking of a single diagnostic indicator and cannot independently represent the overall diagnostic performance. Future high-quality, multicenter studies are still needed to validate the clinical utility of different quantitative ultrasound techniques in the screening and staging of MASLD.

Clinical Trial Number: Not applicable.

Keywords: MASLD; MRI-PDFF; Quantitative ultrasound-derived techniques; Diagnostic performance; NMA.

1. Introduction

Metabolic dysfunction-associated steatotic liver disease (MASLD), previously termed non-alcoholic fatty liver disease (NAFLD), is a chronic hepatic condition primarily caused by

ARTICLE IN PRESS

metabolic dysfunction. This disease can progress from simple hepatic steatosis to non-alcoholic steatohepatitis (NASH), liver fibrosis, and cirrhosis, ultimately hepatocellular carcinoma (HCC) in some cases. As the chronic liver disease with the highest global prevalence, MASLD affects over 30% of adults worldwide(1), with particularly high prevalence among metabolic high-risk populations like the obese and type 2 diabetes mellitus (T2DM) populations (2). Given the ongoing global increase in obesity, T2DM, and metabolic syndrome, MASLD prevalence continues to rise, and it is forecast to be the foremost driver of end-stage liver pathology and the requirement for hepatic transplantation in the immediate future(3, 4). MASLD not only causes localized hepatic pathological damage but also notably elevates the HCC risk in comparison to the healthy population. Patients with MASLD are also more likely to develop multiple complications, including cardiovascular disease, sarcopenia, and extrahepatic tumors (1, 5). Cardiovascular disease represents the primary cause of MASLD-related death(5), while the development of liver fibrosis sharply worsens long-term prognosis, exponentially increasing the risks of cirrhosis and HCC(3). From a health economics peRSpective, the diagnosis, treatment, management of complications, and long-term follow-up of MASLD consume substantial medical resources, imposing a heavy economic burden on patients, families, and healthcare systems globally(3, 4). Since early-stage MASLD (simple hepatic steatosis) can be reversed through lifestyle index tests, whereas treatment becomes markedly more difficult after progression to NASH or fibrosis, and there are no specific pharmacologic therapies(6), early and precise graded diagnosis is critical for improving patient outcomes and reducing the overall disease burden, representing significant clinical and public health value.

Currently, clinical grading of MASLD primarily relies on invasive pathological examinations and non-invasive imaging modalities, as recommended by authoritative guidelines such as those

ARTICLE IN PRESS

from the World Federation for Ultrasound in Medicine and Biology (WFUMB) and the European Association for the Study of the Liver (EASL). Both organizations prioritize non-invasive imaging methods (e.g., magnetic resonance imaging-derived proton density fat fraction (MRI-PDFF), quantitative ultrasound, and conventional B-mode ultrasound) over invasive liver biopsy (LB), though they differ in the preferred order: WFUMB prioritizes MRI-PDFF over quantitative ultrasound, whereas EASL favors quantitative ultrasound over MRI-PDFF(7, 8). LB continues to be the gold standard for MASLD lesion grading, enabling precise hepatic steatosis, inflammation, and fibrosis evaluation, with its diagnostic validity well established in many clinical studies(9). MRI-PDFF, recommended by WFUMB as the gold standard for non-invasive liver fat quantification, demonstrates strong correlation with histological grading and excellent reproducibility and accuracy(8, 10). Conventional B-mode ultrasound is a recommended first-line screening approach due to its convenience, lack of ionizing radiation, and low cost (7, 8). However, all current diagnostic approaches have notable limitations. LB is invasive, carries risks such as pain and bleeding, and is susceptible to sampling errors and interobserver variability, limiting its suitability for routine screening or longitudinal monitoring(9, 11). MRI-PDFF, while highly precise, is costly and has limited accessibility, hindering widespread use in primary care and population screening, and it is not ideal for long-term follow-up(7, 9, 12). Conventional B-mode ultrasound is highly operator-dependent and subjective, with poor sensitivity (SEN) for mild or moderate steatosis (liver fat content < 20%), limiting its ability to provide accurate lesion grading(13, 14). In this context, quantitative ultrasound-derived techniques have rapidly advanced due to their non-invasiveness, accessibility in primary care, objectivity, and reproducibility, offering promising tools for assessing liver fat content and grading MASLD. These techniques include ultrasound-derived fat fraction (UDFF),

ARTICLE IN PRESS

attenuation coefficient (AC), controlled attenuation parameter (CAP), tissue structure index (TSI), tissue attenuation imaging (TAI), speed of sound (SS), and hepatorenal ratio (HRR).

While Gu et al. (10) and Yokoo et al. (15) established MRI-PDFF as the gold standard in their meta-analyses, and Cao et al. (16), Sarah et al. (17), and Jong et al. (18) demonstrated the efficacy of ultrasound techniques like UDFF, CAP, and AC, no head-to-head comparisons and existing study heterogeneity(17-19) remain significant issues. Therefore, this study systematically evaluates and compares the diagnostic performance of UDFF, CAP, AC, TSI, TAI, SS, and HRR for MASLD grading via a network meta-analysis (NMA). Direct and indirect evidence were integrated to identify optimal application scenarios for each technique and provide high-level, evidence-based guidance for clinical practice. This research ultimately seeks to standardize non-invasive diagnosis, improve early detection rates, and enhance MASLD patient prognosis through precise, stage-tailored management.

ARTICLE IN PRESS

2. Materials and Methods

2.1 Research registration

This study followed the PRISMA guidelines(20) and its extension for NMA. The protocol has been registered in PROSPERO (Registration No.: CRD420251250779).

2.2 Search strategy

PubMed, Embase, the Cochrane Library, and Web of Science were searched until December 7, 2025, utilizing Medical Subject Headings (MeSH) and free-text terms: “Non-alcoholic Fatty Liver Disease”, “Sensitivity and Specificity,” and “ROC”. The strategy is provided in Supplementary Table S1. The references of systematic reviews were also checked to ensure a thorough literature search.

ARTICLE IN PRESS

2.3 Eligibility criteria

Inclusion criteria were: (1) Population: Adult patients (≥18) diagnosed with or suspected of having MASLD (or NAFLD); (2) index tests: Ultrasound-derived non-invasive diagnostic techniques, including UDFF, AC, HRR, CAP, ultrasound attenuation parameter (UAP), novel TAI, TSI, and similar techniques; (3) Reference standard: MRI-PDFF as the reference standard (MRI-PDFF is widely recognized as one of the golden standards for the diagnosis and quantification of hepatic steatosis in patients with MASLD/NAFLD because of its strong correlation with histopathological findings. Furthermore, all studies included in our analysis used MRI-PDFF as the reference standard, ensuring methodological consistency across the network. Compared with LB, MRI-PDFF is non-invasive, enables whole-liver assessment, and provides objective quantification of hepatic fat content, making it particularly well aligned to evaluate non-invasive ultrasound-derived diagnostic techniques.); (4) Study type: Diagnostic accuracy studies; and (5) Outcome measures: The primary outcomes were diagnostic accuracy indicators for various ultrasound techniques in differentiating grades of hepatic steatosis (e.g., S1 (mild):≥S1 vs S0, S2 (moderate):≥S2 vs S0-S1, S3 (severe):≥S3 vs S0-S2), including SEN and specificity (SPE).

Exclusion criteria were: (1) Animal studies, case reports, reviews, meta-analyses, comments, letters, editorials, communications, guidelines, commentaries, and conference papers; (2) missing research data or serious errors (For instance, the textual descriptions were substantially inconsistent with the corresponding figures, making accurate data extraction unreliable. In addition, the study did not report the diagnostic accuracy data required for quantitative synthesis, including SEN and SPE for detecting specific grades of hepatic steatosis, the numbers of patients with and without steatosis, or the diagnostic thresholds (grading criteria) used to define each

ARTICLE IN PRESS

steatosis stage.); (3) Duplicates; and (4) unavailable full texts.

2.4 Data extraction

The retrieved literature was uploaded to EndNote. Title and abstract checking was independently carried out by two researchers per the eligibility criteria and followed by full-text reading. Disagreements were addressed by discussing or consulting a third researcher. Using Excel 2016, the same two researchers independently collected first author, publication year, country, design, grouping, grading standards, index test and control measures, baseline characteristics of the study population, and outcome measures.

2.5 Quality assessment

Study quality was rated employing the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) (21) across four domains: patient selection, index test, reference standard, and flow and timing. Each domain was classified as having a “low risk”, “high risk”, or “some concerns”. Dissents were addressed through discussing or consulting a third researcher. Evaluation results are shown as a risk of bias (RoB) graph.

ARTICLE IN PRESS

2.6 Statistical analysis

Diagnostic accuracy networks were constructed for S1, S2, and S3 hepatic steatosis, respectively. The relationships between different diagnostic protocols are presented in the form of network plots. Based on the number of diseased individuals, non-diseased individuals, SEN, and SPE reported in each study, true-positive, false-negative, true-negative, and false-positive data were extracted or reconstructed to form 2×2 contingency tables. The primary analysis employed a Bayesian arm-based bivariate random-effects network meta-analysis model for diagnostic accuracy, implemented using the RStan package in R 4.5(20). The model used the number of true

ARTICLE IN PRESS

positives in the diseased population and the number of true negatives in the non-diseased population as binomial outcomes, jointly modeling SEN and SPE on the logit scale, and accounted for the inherent correlation between SEN and SPE within the same study through correlated study-level random effects. The model incorporated various quantitative ultrasound techniques as fixed effects, while also incorporating between-study heterogeneity and random variation between different diagnostic techniques within the same study. The model was estimated using three Markov chain Monte Carlo chains, each comprising 2,000 burn-in iterations and 10,000 sampling iterations with a thinning interval of 5. Convergence and sampling stability of the model were assessed via trace plots, R-hat, effective sample size, divergent transitions, maximum tree depth, and Bayesian fraction of missing information. An R-hat < 1.10 along with well-mixed chains in the trace plots was considered indicative of basic model convergence. All prior distributions and the complete RStan code are are presented as the original data in the relevant files. Upon completion of model fitting, the pooled SEN, pooled SPE, and their 95% credible intervals for each diagnostic technique were estimated based on the posterior distribution. The DOR was calculated from the jointly estimated SEN and SPE, and the superiority index (SI) was used to comprehensively evaluate the overall diagnostic performance of different techniques. Because the SI simultaneously considers both SEN and SPE, it was adopted as the ranking metric for overall diagnostic performance. As a supplementary and exploratory analysis, univariate Bayesian network meta-analyses were performed for SEN and SPE separately using gemtc. Consistency between the consistency and inconsistency models was evaluated by comparing their Deviance Information Criterion (DIC) values. An absolute difference in DIC of <5 was considered indicative of good global consistency, in which case the more parsimonious consistency model was preferred(21,22). Conversely, a DIC difference of ≥5

ARTICLE IN PRESS

suggested the presence of substantial inconsistency. Therefore, node-splitting analysis was performed to identify locally inconsistent comparisons. For the diseased population, true positives were treated as events to calculate the relative sensitivity (RSe) between different diagnostic techniques; for the non-diseased population, true negatives were treated as events to calculate the relative specificity (RSp) between different diagnostic techniques. Relative sensitivity and relative specificity were reported as RSe and RSp along with their 95% credible intervals (CrIs), respectively. SUCRA was calculated based on the posterior ranking probabilities of SEN and SPE (23-25), respectively. The relationships between different diagnostic protocols are presented in the form of network plots. Possible publication bias was examined through comparison-adjusted funnel plots (26, 27). Integrating findings from the node-splitting approach, cumulative probability ranking plots were generated to visually illustrate the comparative efficacy and ranking distributions of the various diagnostic strategies. SEN-SUCRA reflects only the probability of a given technique ranking high in terms of SEN, whereas SPE-SUCRA reflects only its probability of ranking high in terms of SPE (28, 29). Neither represents the overall diagnostic performance that simultaneously considers both SEN and SPE, which is primarily determined based on the SI, DOR, pooled SEN, and pooled SPE obtained from the RStan bivariate model. These metric-specific rankings cannot simultaneously reflect the trade-off between sensitivity and specificity, and the superiority index derived from the bivariate model is used to summarize overall diagnostic performance.

The RStan bivariate model and the gemtc univariate model have different statistical estimation objectives. The former is used to jointly estimate SEN, SPE, and their correlation, yielding pooled diagnostic accuracy metrics; the latter is used to explore the relative comparisons and rankings of different techniques regarding SEN or SPE individually. Therefore, the results of the

ARTICLE IN PRESS

two sets of models are not required to be completely consistent, and the gemtc results serve solely as supplementary exploratory findings rather than the primary basis for judging overall diagnostic superiority.

3. Results

3.1 Literature search and screening

543 articles were identified. 244 duplicates were removed. Following a preliminary title and abstract checking, 267 articles were excluded. The rest underwent full-text review per the eligibility criteria. 17 articles were ultimately encompassed. The process is provided in Figure 1.

3.2 Basic characteristics

The basic characteristics of eligible studies are provided in Table 1. 17 encompassed studies (30-46)originated from 10 countries (China, Italy, USA, South Korea, Japan, Germany, Greece, Hungary, Turkey, and Canada), consisting of 10 cohort studies and seven cross-sectional studies. All studies applied PDFF values as the grading standard, dividing participants into patient and non-patient cohorts (S0). The former was further split: S1 (mild), S2 (moderate), and S3 (severe). There were 1,904 MASLD patients (913 with S1, 649 with S2, and 342 with S3) and 1,041 non-patients. For the total study population, the male-to-female ratio was 1,804/1,141. Their mean age was 49.11 ± 14.69, and the mean BMI was 27.72 ± 6.04 kg/m². All cohorts included in this study exhibited a broad range of body mass index (BMI), reflecting genuine clinical heterogeneity. The mean BMI across the included studies ranged from 26.3 to 32.9 kg/m², with a substantial proportion of participants classified as overweight (BMI ≥25 kg/m²) or obese (BMI ≥30 kg/m²). Outcome information is provided in Supplementary Table S2. Detailed information on the index tests, including the manufacturer and model of the ultrasound system, probe type and frequency, and image-processing algorithms, is provided in Supplementary Table S3.

ARTICLE IN PRESS

3.3 Inclusion of methodological quality assessment results

In patient selection, all studies displayed a low RoB because the inclusion and exclusion criteria for study participants were clearly defined, indicating no evident selection bias. Regarding the index test, since the results of quantitative ultrasound-derived techniques are objective indicators, the interpretation of the index test results was conducted with knowledge of the gold standard results; therefore, all studies were assessed as having an unclear RoB. Regarding the reference standard, all studies adopted MRI-PDFF as the gold standard, which demonstrates high diagnostic accuracy. The reference standard results were interpreted without knowing the index test results; thus, all studies exhibited low RoB. In flow and timing, all studies showed a generally low RoB. 16 studies exhibited unclear RoB because not all data from the recruited participants were encompassed, while one study displayed a low RoB because all recruited participant data were encompassed. Overall, eligible studies demonstrated a moderate-low RoB (Figure 2).

3.4 NMA results

3.4.1 Network diagram

The 17 included studies involved 17 different index tests: AC, AC+TAI, CAP, HRR, NLV, radiofrequency signal-based fat measurement (RFM)+attenuation coefficient estimation (ACE)+fat index (FI), RFM+ACE+harmonic imaging (HI), SC+TSI, SMI+VI, SS, SWA, TAI, TSI, UAP, UDFF, USFF and Deep-USFF. The network of competing index tests is depicted in Figure 3. Here, the width of each edge is weighted by the number of available studies for that pairwise comparison, and the size of each node is scaled according to the total number of participants receiving that index test. To evaluate local inconsistencies, node-splitting analysis was performed on all outcomes with closed loops. Most p-values were greater than 0.05,

ARTICLE IN PRESS

suggesting good overall consistency. Results are detailed in Supplementary Table S4. The number of studies evaluating each ultrasound-derived technique, the corresponding sample sizes, study phases, and detailed characteristics of the index tests, including the manufacturer and model of the ultrasound system, image-processing algorithm, probe type, and probe frequency, are summarized in Supplementary Table S3.

3.4.2 Consistency Level

Network consistency was assessed by comparing the DIC between the consistency and inconsistency models. The differences in DIC (ΔDIC) for S1 SEN, S2 SEN, and S2 SPE were all below the predefined threshold of 5, indicating acceptable global consistency (S1 SEN: ΔDIC=0.406; S2 SEN: ΔDIC=1.913; S2 SPE: ΔDIC=0.540). In contrast, the ΔDIC values for S1 SPE, S3 SEN, and S3 SPE exceeded the threshold (S1 SPE: ΔDIC=5.207; S3 SEN: ΔDIC=6.547; S3 SPE: ΔDIC=10.553), suggesting substantial inconsistency within these networks (Supplementary Appendix 3). Accordingly, the consistency model was adopted for the analyses of S1 SEN, S2 SEN, and S2 SPE, whereas node-splitting analyses were subsequently performed for S1 SPE, S3 SEN, and S3 SPE to identify potential sources of local inconsistency.

3.4.3 Results of the RStan Bivariate Diagnostic Accuracy NMA

The arm-based bivariate random-effects network meta-analysis of diagnostic accuracy constructed using RStan jointly estimated the sensitivity (SEN) and specificity (SPE) of different diagnostic tools within the same model, with the results presented in Supplementary Table S5.

The superiority index (SI), diagnostic odds ratio (DOR), pooled SEN, and pooled SPE for each ultrasound-derived diagnostic technique are presented in Supplementary Table S5.

Bayesian NMA of diagnostic accuracy demonstrated excellent model convergence, with all potential scale reduction factors (R-hat) equal to 1.00. Diagnostic performance varied across the

ARTICLE IN PRESS

evaluated techniques. For S1 steatosis, SWA achieved the highest pooled SEN (0.96, 95% CrI: 0.73-1.00) and the highest DOR (373,719.55, 95% CrI: 8.08-853,672.04), whereas RFM-ACE-HI demonstrated the highest pooled SPE (0.95, 95% CrI: 0.79-1.00) together with a high superiority index (SI=14.25). Notably, the DOR estimate for SWA was associated with a very wide CrI, likely attributable to zero-event studies included in the analysis, indicating considerable uncertainty in the estimated diagnostic performance. Therefore, this finding should be interpreted cautiously.

For S2 steatosis, UDFF yielded the highest pooled SEN (0.92, 95% CrI: 0.85-0.96), whereas SWA showed the highest DOR (234.01, 95% CrI: 1.86-1,077.23). RFM-ACE-HI achieved the highest pooled SPE (0.91, 95% CrI: 0.71-0.99) and the highest superiority index (SI=15.72). Similar to the findings for S1 steatosis, the wide CrI of the DOR for SWA suggests substantial uncertainty, probably resulting from zero-event studies, and therefore warrants cautious interpretation.

For S3 steatosis, TAI demonstrated the highest pooled SEN (0.98, 95% CrI: 0.88-1.00) and the highest DOR (2,498,255.36, 95% CrI: 34.43-1,309,055.51), whereas both UDFF (0.89, 95% CrI: 0.83-0.93) and Deep-USFF (0.89, 95% CrI: 0.68-0.97) achieved the highest pooled SPE. In addition, Deep-USFF ranked highly in overall diagnostic performance (SI=14.61). As with SWA, the DOR estimate for TAI was accompanied by a wide CrI, probably due to zero-event studies, indicating considerable uncertainty in the estimated effect size. Accordingly, these findings should be interpreted with caution.

ARTICLE IN PRESS

ARTICLE IN PRESS

3.4.4 Exploratory Univariate Network Comparisons of RSe and RSp

To supplementally evaluate the relative performance of different diagnostic tools in single diagnostic metrics, a univariate Bayesian network meta-analysis was further performed for SEN and SPE separately using the gemtc package. Because this analysis did not jointly model SEN and SPE, its results include only RSe, RSp, and SUCRA calculated based on SEN and SPE individually. Consequently, the results in this section serve solely as exploratory single-metric comparisons.

3.4.4.1 Diagnostic efficacy for S1 grading

In terms of diagnostic SEN for S1, 17 studies reported findings showing several significant advantages (Table 2). Compared with CAP, UAP (RSe=1.15, 95% CrIs: 1.00-1.33), NLV (RSe=1.13, 95% CrIs: 1.00-1.28) and UDFF (RSe=1.12, 95% CrIs: 1.04-1.22) showed notably better performance. Relative to AC, SWA (RSe=1.30, 95% CrIs: 1.01-1.66), UAP (RSe=1.24, 95% CrIs: 1.04-1.48), NLV (RSe=1.22, 95% CrIs: 1.06-1.40), UDFF (RSe=1.21, 95% CrIs: 1.05-1.39) and SMI-VI (RSe=1.21, 95% CrIs: 1.04-1.39) showed notably better performance. Relative to HRSe, SWA (RSe=1.33, 95% CrIs: 1.02-1.72), UAP (RSe=1.27, 95% CrIs: 1.04-1.54), NLV (RSe=1.24, 95% CrIs: 1.04-1.49), UDFF (RSe=1.24, 95% CrIs: 1.06-1.45) and SMI-VI (RSe=1.23, 95% CrIs: 1.03-1.48) showed notably better performance. Other pairwise comparisons were statistically insignificant (P>0.05). According to cumulative probability ranking, SWA (SUCRA: 99.80%), UAP (SUCRA: 78.69%), and RFM-ACE-HI (SUCRA: 72.21%) are likely the top three techniques for improving diagnostic SEN for S1 (Figure 4 and Supplementary Table S6).

Regarding diagnostic SPE for S1, several significant differences were identified (Table 3). Compared with NLV, RFM-ACE-HI (RSp=1.36, 95% CrIs: 1.09-1.69), SC-TSI (RSp=1.35, 95%

ARTICLE IN PRESS

CrIs: 1.07-1.69), AC (RSp=1.30, 95% CrIs: 1.08-1.55), Deep-USFF (RSp=1.26, 95% CrIs: 1.00-1.59) and UDFF (RSp=1.21, 95% CrIs: 1.01-1.45) showed notably better performance. Relative to AC, RFM-ACE-HI (RSp=1.42, 95% CrIs: 1.12-1.79), SC-TSI (RSp=1.41, 95% CrIs: 1.08-1.85), AC (RSp=1.36, 95% CrIs: 1.05-1.75), Deep-USFF (RSp=1.32, 95% CrIs: 1.01-1.74) and UDFF (RSp=1.27, 95% CrIs: 1.00-1.60), showed notably better performance. Other pairwise comparisons were statistically insignificant (P>0.05). According to the cumulative probability ranking, RFM-ACE-HI (SUCRA: 88.15%), USFF (SUCRA: 75.42%), and Deep-USFF (SUCRA: 71.60%) are likely the top three techniques for improving diagnostic SPE for S1 (Figure 4 and Supplementary Table S6).

3.4.4.2 Diagnostic efficacy for S2 grading

15 studies reported diagnostic SEN for S2 (Table 4). Compared with CAP, USFF (RSe=1.09, 95% CrIs: 1.01-1.17) showed notably better performance. Relative to TAI, USFF (RSe=1.17, 95% CrIs: 1.04-1.31), TSI (RSe=1.14, 95% CrIs: 1.01-1.29) and UDFF (RSe=1.14, 95% CrIs: 1.02-1.28) showed notably better performance. Relative to AC-TAI, 13 techniques, including USFF (RSe=1.46, 95% CrIs: 1.20-1.78), TSI (RSe=1.43, 95% CrIs: 1.17-1.75), and UDFF (RSe=1.43, 95% CrIs: 1.17-1.73), showed notably better performance. Other pairwise comparisons were statistically insignificant (P>0.05). According to cumulative probability ranking, USFF (SUCRA: 84.25%), UDFF (SUCRA: 81.45%), and RFM-ACE-HI (SUCRA: 79.29%) are likely the top three techniques for improving diagnostic SEN for S2 (Figure 4 and Supplementary Table S6).

Several significant differences were observed in diagnostic SPE for S2 (Table 5). Compared with TSI, TAI (RSp=1.19, 95% CrIs: 1.00-1.42) showed notably better performance. Relative to CAP, 3 techniques, including Deep-USFF (RSp=1.34, 95% CrIs: 1.01-1.80), TAI (RSp=1.29, 95% CrIs:

ARTICLE IN PRESS

1.04-1.20), and AC (RSp=1.28, 95% CrIs: 1.02-1.59), showed notably better performance. Relative to NLV, TAI (RSp=1.31, 95% CrIs: 1.00-1.71) and AC (RSp=1.30, 95% CrIs: 1.04-1.61) showed notably better performance. Other pairwise comparisons were statistically insignificant (P>0.05). By cumulative probability ranking, RFM-ACE-HI (SUCRA: 81.43%), RFM-ACE-FI (SUCRA: 80.37%), and USFF (SUCRA: 77.40%) are likely the top three techniques for improving diagnostic SPE for S2 (Figure 4 and Supplementary Table S6).

3.4.4.3 Diagnostic efficacy for S3 grading

13 studies reported results on diagnostic SEN for S3 (Table 6). Compared with UDFF, TAI (RSe=1.19, 95% CrIs: 1.02-1.38), SS (RSe=1.16, 95% CrIs: 1.00-1.35), SMI-VI (RSe=1.14, 95% CrIs: 1.00-1.31) and AC (RSe=1.13, 95% CrIs: 1.00-1.28) showed notably better performance. Relative to CAP, TAI (RSe=1.23, 95% CrIs: 1.03-1.47), SS (RSe=1.20, 95% CrIs: 1.00-1.44), SMI-VI (RSe=1.18, 95% CrIs: 1.00-1.40) and AC (RSe=1.17, 95% CrIs: 1.00-1.37) showed notably better performance. Other pairwise comparisons were statistically insignificant (P>0.05). By cumulative probability ranking, TAI (SUCRA: 96.77%), SS (SUCRA: 81.49%), and USFF (SUCRA: 68.82%) are likely the top three techniques for improving diagnostic SEN for S3 (Figure 4 and Supplementary Table S6).

Several significant differences were observed in diagnostic SPE for S3 (Table 7). Compared with AC, UDFF (RSp=1.45, 95% CrIs: 1.02-2.06) showed notably better performance. Compared with NLV, 6 techniques, including Deep-USFF (RSp=2.03, 95% CrIs: 1.13-3.66), UDFF (RSp=1.90, 95% CrIs: 1.27-2.85), and SWA (RSp=1.97, 95% CrIs: 1.05-3.68), showed notably better performance. Relative to SS, 5 techniques, including Deep-USFF (RSp=2.34, 95% CrIs: 1.16-4.72), UDFF (RSp=2.19, 95% CrIs: 1.26-3.83), and SWA (RSp=2.26, 95% CrIs: 1.08-4.73), showed notably better performance. Other pairwise comparisons were statistically insignificant

ARTICLE IN PRESS

(P>0.05). By cumulative probability ranking, SWA (SUCRA: 80.64%), Deep-UDFF (SUCRA: 76.91%), and UDFF (SUCRA: 73.68%) are likely the top three techniques for improving diagnostic SPE for S3 (Figure 4 and Supplementary Table S6).

3.4.5 Publication bias

Publication bias was assessed using funnel plots in conjunction with Egger's regression test (Figure 5). For the diagnostic performance of S2 and S3 steatosis grading, the funnel plots appeared largely symmetrical, and Egger's tests were not statistically significant (all P>0.05; specifically, S2 **SEN**: P=0.2324; S2 **SPE**: P=0.0600; S3 **SEN**: P=0.2186; and S3 **SPE**: P=0.4009), suggesting no evidensce of significant publication bias. For S1 steatosis, although the funnel plot also appeared approximately symmetrical, Egger's test indicated significant publication bias for **SEN** (P=0.0067), whereas no significant publication bias was detected for **SPE** (P=0.1226). Therefore, publication bias cannot be excluded for the **SEN** analysis of S1 steatosis. The number of studies available for each outcome was relatively limited. Consequently, the statistical power of the funnel plots and Egger's tests is inherently limited in reliably detecting asymmetry or true publication bias.

ARTICLE IN PRESS

4. Discussion

This NMA evaluated the comparative diagnostic performance of quantitative ultrasound-derived techniques for grading hepatic steatosis using MRI-PDFF as the reference standard. The bivariate random-effects model demonstrated that SWA, UDFF, and TAI achieved the highest pooled SEN for detecting S1, S2, and S3 steatosis, respectively, whereas RFM+ACE+HI showed superior SPE for S1 and S2 steatosis, and UDFF/Deep-USFF demonstrated the highest SPE for S3 steatosis. Furthermore, SUCRA rankings derived from gemtc consistently suggested that SWA, UDFF, and TAI ranked highly for SEN, while SWA and RFM+ACE+HI were among the

ARTICLE IN PRESS

top-performing techniques for SPE. Nevertheless, because the bivariate random-effects model and the SUCRA-based network ranking evaluate diagnostic performance from different methodological peRSpectives, discrepancies between pooled diagnostic estimates and ranking probabilities are expected. Therefore, SUCRA rankings should be interpreted as exploratory evidence of relative comparative performance rather than definitive evidence of clinical superiority. The core necessity of this study arises from the clinical challenge of technology selection. While conventional meta-analyses have confirmed the effectiveness of individual techniques, the absence of direct head-to-head comparisons makes it difficult for clinicians to identify the most appropriate diagnostic tool for specific steatosis grades. The contribution of this study lies in synthesizing available direct and indirect evidence from 17 international studies to systematically rank 15 quantitative ultrasound techniques within a unified Bayesian framework. This approach helps bridge small-sample gaps in individual studies, though the findings should be viewed as exploratory guidance rather than definitive clinical rules. From a clinical peRSpective, our findings provide comparative evidence to support the selection of quantitative ultrasound techniques according to specific diagnostic objectives while acknowledging their respective technical limitations.

In the bivariate random effects model, SWA is recommended as the preferred technique for improving the diagnostic SEN of S1 grading. This may be because SWA not only measures acoustic attenuation but also reflects the viscoelastic properties of liver tissue. Hepatic steatosis alters the viscous characteristics of liver tissue, and SWA can capture these microstructural changes by detecting shear wave attenuation. In this study, SWA was calculated across a wide frequency range (0-1000 Hz) using an improved frequency-shift method. Compared with single-frequency measurements, this full-band analysis may more sensitively detect subtle

ARTICLE IN PRESS

acoustic changes caused by early fat accumulation [43]. RFM+ACE+HI is recommended as the preferred technique for improving the diagnostic SPE of S1 and S2 grading and for comprehensively enhancing diagnostic efficacy for these stages. This advantage may be attributed to the use of harmonic imaging. Compared with fundamental imaging, harmonic imaging can significantly suppress reverberation artifacts originating from the body wall, particularly from subcutaneous fat. In obese patients, conventional ultrasound measurements are often affected by such artifacts, leading to measurement errors or false-positive results. By filtering out these interferences, RFM+ACE+HI can provide a purer attenuation signal and thus improve diagnostic SPE [44]. UDFF demonstrated the best overall performance for improving SEN in detecting moderate steatosis (S2) and SPE in identifying severe steatosis (S3). This advantage is likely attributable to its unique technical framework and multiparametric quantitative design. By integrating quantitative ultrasound biomarkers, particularly AC and backscatter coefficient (BSC), UDFF employs an optimized mathematical model to directly quantify hepatic fat content, thereby reducing the subjectivity inherent in conventional B-mode ultrasonography. In addition, previous multicenter and prospective validation studies have shown that UDFF maintains high SEN for detecting moderate steatosis while preserving excellent SPE in severe steatosis. This performance is probably related to its advanced radiofrequency signal processing and algorithmic calibration, which mitigate signal saturation and distortion at high hepatic fat fractions, thereby enabling more accurate grading across a broad spectrum of disease severity. Consequently, UDFF represents a promising non-invasive imaging biomarker for accurate staging and longitudinal monitoring of hepatic steatosis in patients with MASLD. [30][38] Deep-USFF demonstrated the highest SPE for S3 steatosis. This superior performance is likely attributable to the complementary advantages of multiparametric imaging combined with deep

ARTICLE IN PRESS

learning-based feature extraction. Specifically, the model integrates conventional B-mode ultrasound with TAI and TSI, enabling comprehensive characterization of hepatic tissue microstructure while overcoming the qualitative and operator-dependent limitations of conventional ultrasound. Furthermore, the 2D-CNN automatically extracts high-dimensional spatial and textural features that cannot be captured by conventional single-parameter analyses. This deep feature representation improves robustness against image noise and enhances discrimination of severe hepatic steatosis, thereby contributing to the high SPE observed for S3 disease. [36]

TAI shows the best performance in improving the diagnostic SPE of S3 grading, likely because it measures the energy loss of ultrasonic waves during propagation through tissue. Studies have demonstrated that as hepatic fat content increases, particularly when progressing to S3, acoustic attenuation shows a significant linear increase. Compared with TSI, which is based on scattering distribution, scattering signals may change or saturate in severe steatosis. In contrast, TAI can maintain stable differentiation even with increasing imaging depth and extremely high fat content, resulting in more robust diagnostic performance for severe steatosis grading [46].

Regarding patient BMI distribution, body composition and BMI represent important confounding factors affecting the accuracy of quantitative ultrasound. Multiple included studies, including the multivariate regression analysis by Hayashi et al. [35], have demonstrated independent associations between ultrasound parameters, such as SS and AC, and patient BMI. As BMI increases, or as subcutaneous fat thickness and the skin-to-liver capsule distance (SCD) become greater, ultrasound waves undergo increased energy attenuation and scattering while penetrating the thicker abdominal wall. This interferes with parameters derived from echo signal attenuation (e.g., CAP or AC), thereby reducing the concordance (ICC) between quantitative

ARTICLE IN PRESS

ultrasound and the reference standard, MRI-PDFF. To address this issue, some advanced quantitative ultrasound systems have introduced deep abdominal (DAX) or XL probes, or adopted harmonic imaging modes (e.g., RFM-ACE-HI), which more effectively suppress abdominal wall reverberation artifacts than fundamental imaging, thereby maintaining good diagnostic accuracy in individuals with high BMI. These findings suggest that BMI heterogeneity should be carefully considered in clinical practice and during result interpretation. When appropriate, dedicated probes or complementary non-invasive indices (e.g., HSI) may be used to facilitate stratified calibration in overweight or obese individuals.

In terms of operator experience, the acquisition of quantitative ultrasound parameters is highly dependent on standardized image-plane optimization, appropriate region-of-interest (ROI) selection (e.g., avoiding large blood vessels and rib shadowing while maintaining a sampling depth of approximately 1.5 cm below the liver capsule), and patient respiratory cooperation. Operator proficiency directly influences the quality of the acquired RF data. The prospective and multicenter cohorts included in this study all adopted rigorous quality-control protocols, and most examinations were performed by experienced physicians specializing in abdominal and quantitative ultrasound. Although some studies did not report the operators' exact years of experience, all stated that examinations were conducted by certified medical personnel. Furthermore, several included studies, including those by Wang et al. [30] and De Robertis et al. [31], demonstrated that standardized multi-site sampling combined with blinded image assessment achieved excellent intra- and interobserver reproducibility (intraclass correlation coefficient [ICC] $\geq$ 0.86-0.96), indicating that standardized training and operator experience are fundamental to ensuring the diagnostic performance of quantitative ultrasound.

The limitations of this study primarily relate to the study methodology, the quality of the

ARTICLE IN PRESS

included literature, and the inherent structural constraints of the NMA evidence base. Specifically, only 17 studies evaluating 17 distinct index tests were included, resulting in a limited evidence base for each ultrasound technique. Accordingly, the interpretation of the SUCRA rankings should be tempered by the methodological limitations of the evidence network. Many of the evaluated quantitative ultrasound techniques were represented by only a single study or a very small number of studies, resulting in a sparse network in which most comparative estimates relied heavily on indirect evidence derived from relatively small sample sizes. This structural sparsity, together with the asymmetry and small-study effects observed in the funnel plots (particularly for S1 SEN, P=0.0067), increases the susceptibility of the results to publication bias. In sparse networks characterized by small-study effects, the preferential publication of small studies reporting exaggerated positive findings may disproportionately influence pooled relative effect estimates. Consequently, techniques ranked highest by SUCRA (such as SWA for S1 SEN and TAI for S3 SEN) may have had their overall diagnostic performance overestimated. Therefore, although SUCRA rankings provide a useful comparative framework, the combined effects of network sparsity, limited sample sizes, and potential publication bias indicate that these rankings should be regarded as exploratory rather than definitive and interpreted with appropriate caution. Future studies should expand both the number of searched databases and the scope of the literature search to identify additional eligible studies, thereby improving the comprehensiveness of diagnostic performance evaluations and strengthening the network of direct comparisons. Another important limitation relates to the reference standard used across the included studies. Although MRI-PDFF is widely accepted as an advanced non-invasive imaging modality, it is not the definitive gold standard; histopathological assessment by LB remains the clinical reference standard. Because MRI-PDFF quantifies hepatic fat content based on proton

ARTICLE IN PRESS

density rather than direct histopathological evaluation, it is subject to inherent measurement limitations and the potential for misclassification. Consequently, evaluating quantitative ultrasound techniques against this non-invasive surrogate rather than histopathology may have influenced the estimated diagnostic performance and the interpretation of our findings. Future diagnostic studies should incorporate histopathological validation whenever feasible to provide more robust evidence. Furthermore, methodological gaps within the primary literature also affected the analysis. Not all encompassing studies incorporated the complete data from recruited participants. In some cases, patients were unable to undergo diagnosis with the relevant techniques, the interval between the index test and the gold standard examination was excessively long, or consent was withdrawn. In other cases, patients were excluded during subsequent screening for failing to meet eligibility criteria. These factors may have influenced the overall conclusions of the analysis. Future research, including a larger number of eligible studies, may help address this limitation. Additionally, due to the limitations of the available data and the small overall evidence pool, detailed subgroup analyses based on variables such as age, BMI, sex, and different technical parameters could not be performed. This may also have influenced the interpretation of the results to some extent. Future studies shall expand research on related technologies and provide more detailed datasets to improve the evaluation of diagnostic efficacy.

ARTICLE IN PRESS

5. Conclusion

This study demonstrates that quantitative ultrasound-related techniques hold significant application potential for the non-invasive diagnosis of varying degrees of hepatic steatosis, yet notable trade-offs exist between sensitivity (SEN) and specificity (SPE) across different techniques. The RStan bivariate diagnostic accuracy network meta-analysis revealed that SWA,

ARTICLE IN PRESS

UDFF, and TAI achieved higher pooled SEN in diagnosing S1, S2, and S3 steatosis, respectively; RFM+ACE+HI achieved higher pooled SPE in diagnosing S1 and S2 steatosis, whereas UDFF and Deep-USFF exhibited higher pooled SPE in diagnosing S3 steatosis. The SUCRA rankings performed separately for SEN and SPE provide supplementary comparisons of individual diagnostic metrics but do not represent overall diagnostic performance. Given that some techniques are supported by only a limited number of studies and that the credible intervals for several DOR and SI values are wide, the current findings should be regarded as exploratory evidence. Future large-sample, multicenter, head-to-head studies utilizing unified diagnostic thresholds are still warranted to validate the clinical diagnostic value of different quantitative ultrasound techniques.

List of abbreviations

MASLD, Metabolic dysfunction-associated steatotic liver disease; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; HCC, hepatocellular carcinoma; T2DM, type 2 diabetes mellitus; WFUMB, World Federation for Ultrasound in Medicine and Biology; EASL, European Association for the Study of the Liver; MRI-PDFF, magnetic resonance imaging-derived proton density fat fraction; LB, liver biopsy; UDFF, ultrasound-derived fat fraction; AC, attenuation coefficient; CAP, controlled attenuation parameter; TSI, tissue structure index; TAI, tissue attenuation imaging; SS, speed of sound; HRR, hepatorenal ratio; NMA, network meta-analysis; MeSH, Medical Subject Headings; UAP, ultrasound attenuation parameter; SPE, specificity; AUC, area under the curve; QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies; RoB, risk of bias; RR, risk ratios; CrIs, credible intervals; SUCRA, surface under the cumulative ranking curve; RFM, radiofrequency signal-based fat measurement; ACE, attenuation coefficient estimation; FI, fat index; HI, harmonic imaging

ARTICLE IN PRESS

Declarations

Ethics approval and consent to participate

No ethical approval was required as this study did not involve human participants or laboratory animals.

Consent for publication

Not applicable.

Consent to Participate

Not applicable.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests.

ARTICLE IN PRESS

Funding Declaration

This work was supported by the Department of Science and Technology of Liaoning province (2024JH2/101900012).

Authors' contributions

Xingya Liu: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - Original Draft, Visualization. Funing Liu: Investigation, Data curation, Writing - Original Draft, Visualization. Xuechao Zhai: Methodology, Software, Formal analysis. Wenrui Liu: Methodology, Writing - Review & Editing, Project administration. Hongbo Wang: Conceptualization, Writing - Review & Editing, Supervision, Resources, Project administration, Funding acquisition, and all authors commented on previous versions of the manuscript. All

ARTICLE IN PRESS

authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Reference

1. Rinella ME, Lazarus JV, Ratziu V, Francque SM, Sanyal AJ, Kanwal F, et al. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. J Hepatol. 2023;79(6):1542-56.

2. Eslam M, Newsome PN, Sarin SK, Anstee QM, Targher G, Romero-Gomez M, et al. A new definition for metabolic dysfunction-associated fatty liver disease: An international expert consensus statement. J Hepatol. 2020;73(1):202-9.

3. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology. 2016;64(1):73-84.

4. Younossi Z, Anstee QM, Marietti M, Hardy T, Henry L, Eslam M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. Nat Rev Gastroenterol Hepatol. 2018;15(1):11-20.

5. Lee HH, Lee HA, Kim EJ, Kim HY, Kim HC, Ahn SH, et al. Metabolic dysfunction-associated steatotic liver disease and risk of cardiovascular disease. Gut. 2024;73(3):533-40.

6. Vilar-Gomez E, Martinez-Perez Y, Calzadilla-Bertot L, Torres-Gonzalez A, Gra-Oramas B, Gonzalez-Fabian L, et al. Weight Loss Through Lifestyle Modification Significantly Reduces Features of Nonalcoholic Steatohepatitis. Gastroenterology. 2015;149(2):367-78.e5; quiz e14-5.

7. Ferraioli G, Berzigotti A, Barr RG, Choi BI, Cui XW, Dong Y, et al. Quantification of Liver

ARTICLE IN PRESS

Fat Content with Ultrasound: A WFUMB Position Paper. Ultrasound Med Biol. 2021;47(10):2803-20.

8. EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. Diabetologia. 2016;59(6):1121-40.

9. Rockey DC, Caldwell SH, Goodman ZD, Nelson RC, Smith AD. Liver biopsy. Hepatology. 2009;49(3):1017-44.

10. Gu J, Liu S, Du S, Zhang Q, Xiao J, Dong Q, et al. Diagnostic value of MRI-PDFF for hepatic steatosis in patients with non-alcoholic fatty liver disease: a meta-analysis. Eur Radiol. 2019;29(7):3564-73.

11. Ferraioli G, Filice C, Castera L, Choi BI, Sporea I, Wilson SR, et al. WFUMB guidelines and recommendations for clinical use of ultrasound elastography: Part 3: liver. Ultrasound Med Biol. 2015;41(5):1161-79.

12. Rinella ME, Neuschwander-Tetri BA, Siddiqui MS, Abdelmalek MF, Caldwell S, Barb D, et al. AASLD Practice Guidance on the clinical assessment and management of nonalcoholic fatty liver disease. Hepatology. 2023;77(5):1797-835.

13. Hernaez R, Lazo M, Bonekamp S, Kamel I, Brancati FL, Guallar E, et al. Diagnostic accuracy and reliability of ultrasonography for the detection of fatty liver: a meta-analysis. Hepatology. 2011;54(3):1082-90.

14. Saadeh S, Younossi ZM, Remer EM, Gramlich T, Ong JP, Hurley M, et al. The utility of radiological imaging in nonalcoholic fatty liver disease. Gastroenterology. 2002;123(3):745-50.

15. Yokoo T, Serai SD, Pirasteh A, Bashir MR, Hamilton G, Hernando D, et al. Linearity, Bias, and Precision of Hepatic Proton Density Fat Fraction Measurements by Using MR Imaging: A Meta-Analysis. Radiology. 2018;286(2):486-98.

ARTICLE IN PRESS

ARTICLE IN PRESS

16. Xinge CAO YZ, Lizhuo JIA, Jianghong CHEN, Yi DONG. Comparison of the diagnostic value of ultrasound-derived fat fraction, controlled attenuation parameter, and hepatic/renal ratio in the grading of hepatic steatosis in metabolic associated fatty liver disease. Journal of Clinical Hepatology. 2025;41(9):1788-94.

17. Verdan S, Torri GB, Marcos VN, Moreira MHS, Defante MLR, Fagundes MDC, et al. Ultrasound-derived fat fraction for diagnosing hepatic steatosis: a systematic review and meta-analysis. Eur Radiol. 2025;35(11):7421-30.

18. Jang JK, Choi SH, Lee JS, Kim SY, Lee SS, Kim KW. Accuracy of the ultrasound attenuation coefficient for the evaluation of hepatic steatosis: a systematic review and meta-analysis of prospective studies. Ultrasonography. 2022;41(1):83-92.

19. Xue LY, Ding H, xU HX. Chinese expert consensus on ultrasound-derived fat fraction for the evaluation of metabolic dysfunction-associated fatty liver disease (2025 edition). J Clin Ultrasound Med. 2025;28(1).

20. Nyaga VN, Aerts M, Arbyn M. ANOVA model for network meta-analysis of diagnostic test accuracy data. Stat Methods Med Res. 2018;27(6):1766–178421.

21. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. J R Stat Soc Ser B Stat Methodol. 2002;64(4):583-639.

22. Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 4: Evidence synthesis in decision making: a Bayesian framework for network meta-analysis. London: National Institute for Health and Care Excellence; 2013.

23. Dias S, Sutton AJ, Welton NJ, Ades AE. Heterogeneity: Subgroups, Meta-Regression, Bias And Bias-Adjustment. NICE Decision Support Unit Technical Support Documents. London: National Institute for Health and Care Excellence; 2012.

ARTICLE IN PRESS

ARTICLE IN PRESS

24. Bois FY. Bayesian inference. Methods Mol Biol. 2013;930:597-636.

25. Hamra G, MacLehose R, Richardson D. Markov chain Monte Carlo: an introduction for epidemiologists. Int J Epidemiol. 2013;42(2):627-34.

26. Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. PLoS One. 2013;8(10):e76654.

27. Whegang Youdom S, Tahar R, Basco LK. Comparison of anti-malarial drugs efficacy in the treatment of uncomplicated malaria in African children and adults using network meta-analysis. Malar J. 2017;16(1):311.

28. Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. BMC Med Res Methodol. 2015;15:58.

29. Trinquart L, Attiche N, Bafeta A, Porcher R, Ravaud P. Uncertainty in Treatment Rankings: Reanalysis of Network Meta-analyses of Randomized Trials. Ann Intern Med. 2016;164(10):666-73.

30. Wang P, Song D, Han J, Zhang J, Chen H, Gao R, et al. Comparing Three Ultrasound-Based Techniques for Diagnosing and Grading Hepatic Steatosis in Metabolic Dysfunction-Associated Steatotic Liver Disease. Acad Radiol. 2025;32(4):1949-57.

31. De Robertis R, Spoto F, Autelitano D, Guagenti D, Olivieri A, Zanutto P, et al. Ultrasound-derived fat fraction for detection of hepatic steatosis and quantification of liver fat content. Radiol Med. 2023;128(10):1174-80.

32. Gao J, Zapata I, Chen J, Erpelding TN, Adamson C, Park D. Quantitative Ultrasound Biomarkers to Assess Nonalcoholic Fatty Liver Disease. J Ultrasound Med. 2023;42(8):1675-88.

33. Gong P, Zhang J, Huang C, Lok UW, Tang S, Liu H, et al. Novel Quantitative Liver Steatosis Assessment Method With Ultrasound Harmonic Imaging. J Ultrasound Med. 2025;44(1):77-85.

ARTICLE IN PRESS

34. Jeon SK, Lee JM, Joo I, Park SJ. Quantitative Ultrasound Radiofrequency Data Analysis for the Assessment of Hepatic Steatosis in Nonalcoholic Fatty Liver Disease Using Magnetic Resonance Imaging Proton Density Fat Fraction as the Reference Standard. Korean J Radiol. 2021;22(7):1077-86.

35. Hayashi M, Sugimoto K, Kakegawa T, Takahashi H, Takeuchi H, Araki Y, et al. Evaluation of attenuation coefficient and sound speed for the diagnosis of liver steatosis: magnetic resonance imaging proton density fat fraction as a reference. J Med Ultrason (2001). 2026;53(1):43-53.

36. Jeon SK, Lee JM, Joo I, Yoon JH, Lee G. Two-dimensional Convolutional Neural Network Using Quantitative US for Noninvasive Assessment of Hepatic Steatosis in NAFLD. Radiology. 2023;307(1):e221510.

37. Jin H, Jiao M, Yu C, Ren T, Chen Q, Dai Z, et al. Discovery of ultrasound-derived fat fraction as a non-invasive tool for MASLD diagnosis. Eur J Med Res. 2025;30(1):1014.

38. Kubale R, Schneider G, Lessenich CPN, Buecker A, Wassenberg S, Torres G, et al. Ultrasound-Derived Fat Fraction for Hepatic Steatosis Assessment: Prospective Study of Agreement With MRI PDFF and Sources of Variability in a Heterogeneous Population. AJR Am J Roentgenol. 2024;222(6):e2330775.

39. Malandris K, Korakas E, Sarakapina A, Kalopitas G, Iatridi F, Liakos A, et al. Accuracy of Controlled Attenuation Parameter for Liver Steatosis in High-Risk Patients for MASLD Using MRI-Proton Density Fat Fraction as Reference Standard. Dig Dis Sci. 2025;70(2):814-24.

40. Mu R, Xia YC, Zhu KY, Lu JY, Luo Q, Zhang L, et al. Diagnostic value of FibroTouch in identifying hepatic steatosis in NAFLD with MRI-PDFF as the reference standard. J Dig Dis. 2023;24(12):691-701.

ARTICLE IN PRESS

ARTICLE IN PRESS

41. Pyo JH, Cho SJ, Jee J, Park G, Ahn J, Kim K, et al. Ethnic Differences in the Diagnostic Accuracy of a Quantitative Ultrasound-Derived Fat Fraction Estimator for Hepatic Steatosis: External Validation in Asian and Caucasian Cohorts. J Clin Ultrasound. 2026;54(1):33-42.

42. Yazdani L, Rafati I, Gesnik M, Nicolet F, Chayer B, Gilbert G, et al. Ultrasound Shear Wave Attenuation Imaging for Grading Liver Steatosis in Volunteers and Patients With Non-alcoholic Fatty Liver Disease: A Pilot Study. Ultrasound Med Biol. 2023;49(10):2264-72.

43. Şendur HN, Özdemir Kalkan D, Cerit MN, Kalkan G, Şendur AB, Özhan Oktar S. Hepatic Fat Quantification With Novel Ultrasound Based Techniques: A Diagnostic Performance Study Using Magnetic Resonance Imaging Proton Density Fat Fraction as Reference Standard. Can Assoc Radiol J. 2023;74(2):362-9.

44. Rónaszéki AD, Budai BK, Csongrády B, Stollmayer R, Hagymási K, Werling K, et al. Tissue attenuation imaging and tissue scatter imaging for quantitative ultrasound evaluation of hepatic steatosis. Medicine (Baltimore). 2022;101(33):e29708.

45. Shao CX, Ye J, Dong Z, Li F, Lin Y, Liao B, et al. Steatosis grading consistency between controlled attenuation parameter and MRI-PDFF in monitoring metabolic associated fatty liver disease. Ther Adv Chronic Dis. 2021;12:20406223211033119.

46. Xue L, Zhu Y, Cheng G, Han H, He N, Chen L, et al. Ultrasound-derived fat fraction for the noninvasive quantification of hepatic steatosis: a prospective multicenter study. Insights Imaging. 2025;16(1):237.

Tables

Table 1. Detailed characteristics of the included studies.

Study ID	Published time	Country	Study design	PDFFvoxel Cutoff-Based Reference Standard	Group	Grade	n	m/f	Age	BMI	Inclusion criteria	PDFF(%)
Wang PP et al. (30)	2024	China	Prospective cohort		MASLD/patient	S1/mild	74	35/39	51.84±12.80	26.34±3.3	PDFF ≥6%	9.90±4.1

ARTICLE IN PRESS

									0		
					S2/moderate	17	4/13	47.82±18.20	28.91±2.87	PDFF ≥17%	18.30±2.80
					S3/severe	24	7/17	41.31±15.72	31.57±4.11	PDFF >22%	26.90±6.60
				non-patient	s0	102	63/39	33.19±12.37	21.02±1.79	PDFF <6%	3.35±1.30
RDe Robertis et al. (31)	2023	Italy	Prospective cross-sectional	MASLD/patient	S1/mild	46	61/61	62±3.90	24.22±1.01	PDFF ≥5%	4.38±2.25
				non-patient	s0	76				PDFF <5%	
				MASLD/patient	S1/mild	43	37/41	50±18	32.3±8.74	PDFF ≥5%	14.17±7.21
					S2/moderate	18				PDFF ≥16.3%	
					S3/severe	17				PDFF ≥21.6%	
Gao J et al. (32)	2023	America	Prospective cohort		s0+F0	22	11/11	58±16	27.6±8.07	PDFF <5%+LS<2.9	3.34±0.74
				non-patient	F1	6	8/17	61±9	32.92±8.08	LS≥2.9	10.39±6.91
					F2	8				LS≥3.5	
					F3	1				LS≥4.0	
					F4	10				LS≥5.0	
					S1/mild	38				PDFF ≥5%	
				MASLD/patient	S2/moderate	13	55/56	56.5±11.8	31.1±6.2	PDFF ≥16.3%	-
					S3/severe	12				PDFF ≥21.7%	
				non-patient	s0	48				PDFF <5%	
				MASLD/patient	S1/mild	23	75/45	49.1±12.6	26.1±3.5	PDFF ≥5%	10.2±7.1
Sun KJ et al.(34)	2021	Korea	Prospective cohort		S2/moderate	59				PDFF ≥10%	

ARTICLE IN PRESS

					non-patient	s0	38			PDFF <5%	
					MASLD/patient	S1/mild	11			PDFF ≥5.2%	
Hayashi M et al.(35)	2025	Japan	Retrospective cross-sectional			S2/moderate	11	26/27	61±15	PDFF ≥11.3%	-
						S3/severe	17			PDFF ≥17.1%	
					non-patient	s0	187			PDFF <5.2%	
					MASLD/patient	S1/mild	79			PDFF ≥5%	
Sun KJ et al. (36)	2023	Korea	Prospective cohort			S2/moderate	37	96/77	51±14	PDFF ≥15%	11.2±7.8
						S3/severe	10			PDFF ≥25%	
					non-patient	s0	47			PDFF <5%	
Jin HR et al. (37)	2025	China	Prospective cross-sectional		MASLD/patient	S1/mild	54	40/14	34.02±2.70	27.27±1.10 PDFF ≥5%	-
					non-patient	s0	49	15/34	29.33±7.45	21.94±0.76 PDFF <5%	
					MASLD/patient	S1/mild	61	112/75	53.8±16.6	29.5±6.8 PDFF ≥6.4%	11.7±10.4
						S2/moderate	15			PDFF ≥17.4%	
				Using PDFFvoxel With Exploratory Cutoffs as the Reference Standard		S3/severe	30			PDFF ≥22.1%	
Reinhard K et al.(38)	2024	Germany	Prospective cohort		non-patient	s0	814			PDFF <6.4%	
						S1/mild	40			PDFF ≥5%	
				Using PDFFvoxel With Exploratory Cutoffs as the Reference Standard	MASLD/patient	S2/moderate	44			PDFF ≥10%	
						S3/severe	35			PDFF ≥20%	
					non-patient	s0	68			PDFF <5%	
Konstantinos M et al. (39)	2025	Greece	Prospective cross-sectional		MASLD/patient	S1/mild	75	66/47	59.6±9.8	32.0±7.8 PDFF ≥5%	9.26±12.58
						S2/moderat	6			PDFF ≥10%	

ARTICLE IN PRESS

					e						
				non-patient	S3/severe	19				PDFF ≥20%	
					s0	31				PDFF <5%	
				MASLD/patient	S1/mild	65	137/33	42.78±12.95	27.25±4.20	PDFF ≥5%	13.34±8.12
					S2/moderate	66				PDFF ≥10%	
					S3/severe	39				PDFF ≥20%	
Mu R et al. (40)	2023	China	Prospective cross-sectional	non-patient	s0	105	62/43	50.46±13.67	23.55±3.27	PDFF <5%	2.90±0.26
				MASLD/patient	S1/mild	129	360/67	55.4±7.1	25.0±2.9	PDFF ≥5%	7.7±6.5
					S2/moderate	168				PDFF ≥10%	
					S3/severe	17				PDFF ≥20%	
Jeung HP et al. (41)	2025	Korea	Retrospective cross-sectional	non-patient	s0	113				PDFF <5%	
				MASLD/patient	S1/mild	17	22/32	54.2±14.0	28.9±2.79	PDFF ≥5%	7.27±1.66
					S2/moderate	37			29.7±3.97	PDFF ≥10%	19.34±7.49
Rónaszéki AD et al.(44)	2022	Hungary	Prospective cohort	non-patient	s0	47	27/20	58.1±10.9	26.4±4.61	PDFF <5%	2.38±1.30
				MASLD/patient	S1/mild	25	59/21	44.3±10.6	28.3±4.7	PDFF ≥5%	-
					S2/moderate	13				PDFF ≥16.3%	
					S3/severe	12				PDFF ≥21.7%	
Halit NS et al. (43)	2023	Turkey	Prospective cohort	non-patient	s0	30	68/34	46.0±13.6	29.1±3.7	PDFF <5%	-
				MASLD/patient	S1/mild	40				PDFF ≥5.5%	
					S2/moderate	37				PDFF ≥14.6%	
					S3/severe	14				PDFF ≥24.3%	
Shao CX et al. (45)	2021	China	Prospective cohort								

ARTICLE IN PRESS

ARTICLE IN PRESS

				non-patient	s0	11				PDFF <5.5%	
				MASLD/patient	S1/mild	10				PDFF ≥6.4%	
					S2/moderate	8				PDFF ≥17.4 %	
Ladan Y et al. (42)	2023	Canada	Prospective cross-sectional		S3/severe	18	25/24	53.0±15.2	27.8±5.4	PDFF >22.1 %	-
				non-patient	s0	13			22.5±1.8	PDFF <6.4%	
Xue LY et al. (46)	2025	China	Prospective cohort	MASLD/patient	S1/mild	151	188/208	41.6±14.1	30.5±7.9	PDFF ≥5.75 %	10.88±9.29

ARTICLE IN PRESS

ARTICLE IN PRESS

Table 2. Relative effects and 95% CrIs (95% credible intervals) of all pairwise panels for SEN based on the Bayesian NMA method (Mild steatosis/ S1).

MRI-PDFF														
1.00 (0.81,1.23)	SWA													
1.05 (0.93,1.18)	1.05 (0.82,1.34)	UAP												
1.07 (0.97,1.17)	1.07 (0.85,1.34)	1.02 (0.87,1.19)	NLV											
1.07 (1.02,1.13)	1.07 (0.86,1.33)	1.02 (0.90,1.17)	1.01 (0.91,1.12)	UDFF										
1.07 (0.97,1.18)	1.07 (0.85,1.35)	1.03 (0.88,1.20)	1.01 (0.90,1.13)	1.00 (0.90,1.12)	SMI-VI									
1.10 (0.97,1.25)	1.10 (0.86,1.40)	1.05 (0.88,1.25)	1.03 (0.88,1.21)	1.02 (0.89,1.17)	1.02 (0.87,1.20)	Deep-USFF								
1.12 (0.95,1.31)	1.12 (0.86,1.45)	1.06 (0.87,1.30)	1.05 (0.87,1.26)	1.04 (0.88,1.23)	1.04 (0.86,1.25)	1.02 (0.83,1.25)	RFM-ACE-FU							
1.15 (0.93,1.41)	1.15 (0.86,1.54)	1.09 (0.86,1.39)	1.08 (0.86,1.35)	1.07 (0.87,1.32)	1.07 (0.85,1.34)	1.04 (0.82,1.33)	1.03 (0.79,1.33)	SC-TSI						
1.15 (1.03,1.28)	1.15 (0.91,1.46)	1.10 (0.93,1.29)	1.08 (0.93,1.25)	1.07 (0.95,1.21)	1.07 (0.92,1.24)	1.05 (0.90,1.21)	1.03 (0.85,1.25)	1.00 (0.79,1.26)	TSI					
1.17 (1.04,1.31)	1.17 (0.92,1.49)	1.12 (0.95,1.32)	1.10 (0.95,1.28)	1.09 (0.96,1.24)	1.09 (0.94,1.27)	1.07 (0.92,1.24)	1.05 (0.86,1.28)	1.02 (0.81,1.29)	1.02 (0.89,1.17)	TAI				
1.17 (1.03,1.33)	1.17 (0.92,1.50)	1.12 (0.94,1.33)	1.10 (0.94,1.29)	1.09 (0.95,1.25)	1.09 (0.93,1.28)	1.07 (0.89,1.28)	1.05 (0.86,1.29)	1.02 (0.80,1.30)	1.02 (0.86,1.21)	1.00 (0.84,1.19)	USFF			
1.20 (1.12,1.30)	1.20 (0.96,1.51)	1.15 (1.00,1.33)	1.13 (1.00,1.28)	1.12 (1.04,1.22)	1.12 (0.99,1.27)	1.10 (0.95,1.27)	1.08 (0.90,1.29)	1.05 (0.85,1.31)	1.05 (0.92,1.20)	1.03 (0.90,1.18)	1.03 (0.89,1.19)	CAP		
1.27 (0.99,1.63)	1.27 (0.92,1.76)	1.21 (0.92,1.60)	1.19 (0.91,1.56)	1.18 (0.92,1.53)	1.18 (0.90,1.54)	1.16 (0.88,1.53)	1.14 (0.85,1.53)	1.11 (0.83,1.49)	1.11 (0.84,1.45)	1.08 (0.82,1.43)	1.08 (0.82,1.43)	1.05 (0.81,1.37)	AC-TAI	
1.26 (1.0	1.26 (0.9	1.20 (0.9	1.18 (0.9	1.18 (0.9	1.17 (0.9	1.15 (0.9	1.13 (0.9	1.10 (0.8	1.10 (0.8	1.08 (0.8	1.08 (0.8	1.05 (0.8	0.99 (0.7	RFM-A

ARTICLE IN PRESS

3,1.54)	5,1.69)	5,1.52)	5,1.48)	6,1.44)	4,1.47)	1,1.46)	0,1.42)	3,1.46)	8,1.38)	6,1.36)	5,1.36)	5,1.30)	2,1.37)	CE-FI		
1.30 (1.14,1.48)	1.30 (1.01,1.66)	1.24 (1.04,1.48)	1.22 (1.06,1.40)	1.21 (1.05,1.39)	1.21 (1.04,1.39)	1.18 (0.98,1.42)	1.16 (0.95,1.43)	1.13 (0.89,1.44)	1.13 (0.95,1.34)	1.11 (0.93,1.32)	1.11 (0.92,1.33)	1.08 (0.93,1.25)	1.02 (0.77,1.35)	1.03 (0.81,1.30)	AC	
1.33 (1.13,1.55)	1.33 (1.02,1.72)	1.27 (1.04,1.54)	1.24 (1.04,1.49)	1.24 (1.06,1.45)	1.23 (1.03,1.48)	1.21 (0.99,1.48)	1.19 (0.95,1.48)	1.16 (0.89,1.50)	1.15 (0.95,1.40)	1.13 (0.93,1.38)	1.13 (0.93,1.38)	1.10 (0.93,1.30)	1.04 (0.78,1.40)	1.05 (0.82,1.35)	1.02 (0.83,1.25)	HRR
1.76 (1.02,3.03)	1.76 (0.98,3.15)	1.68 (0.96,2.93)	1.65 (0.95,2.86)	1.64 (0.95,2.83)	1.64 (0.94,2.84)	1.60 (0.92,2.80)	1.58 (0.90,2.77)	1.54 (0.86,2.74)	1.53 (0.88,2.66)	1.50 (0.86,2.61)	1.50 (0.86,2.62)	1.46 (0.85,2.53)	1.39 (0.76,2.51)	1.39 (0.78,2.48)	1.36 (0.78,2.36)	1.33 (0.76,2.33)	SS

Note: UDFF: Ultrasound-Derived Fat Fraction; CAP: Controlled Attenuation Parameter; HRR: Hepatorenal Ratio; AC: Attenuation Coefficient; NLV: Normalized Local Variance; SMI-VI: Superb Microvascular Imaging Vascular Index; RFM-ACE-FI: Reference Frequency Method-based Attenuation Coefficient Estimation in Fundamental Imaging; RFM-ACE-HI : Reference Frequency Method-based Attenuation Coefficient Estimation in Harmonic Imaging; AC-TAI: Attenuation Coefficient - Tissue Attenuation Imaging; SC-TSI: Scatterer Coefficient - Tissue Scatterer Imaging; SS : Sound Speed; Deep-USFF: Deep-learning based Ultrasound Fat Fraction Estimation; TAI: Tissue Attenuation Imaging; TSI: Tissue Scatterer Imaging; UAP: Ultrasound Attenuation Parameter; USFF: Ultrasound Fat Fraction Estimation; SWA: Shear Wave Attenuation; MRI-PDFF : Magnetic Resonance Imaging - Proton Density Fat Fraction.

Table 3. Relative effects and 95% CrIs of all pairwise panels for SPE based on the Bayesian NMA method (Mild steatosis/ S1).

AC								
1.09 (0.96,1.24)	CAP							
1.03 (0.86,1.23)	0.94 (0.80,1.11)	Deep-USFF						
1.06 (0.90,1.26)	0.98 (0.85,1.12)	1.04 (0.85,1.26)	HRR					
0.94 (0.85,1.04)	0.86 (0.80,0.93)	0.91 (0.79,1.06)	0.88 (0.77,1.00)					
1.30 (1.08,1.55)	1.19 (0.98,1.44)	1.26 (1.00,1.59)	1.22 (0.98,1.51)	NLV				
1.18 (0.94,1.50)	1.09 (0.87,1.36)	1.15 (0.89,1.49)	1.11 (0.87,1.42)	0.91 (0.69,1.20)	AC-TAI			
1.36 (1.05,1.75)	1.25 (0.98,1.59)	1.32 (1.01,1.74)	1.27 (0.98,1.66)	1.05 (0.78,1.40)	1.15 (0.84,1.57)	RFM-ACE-FI		
0.96 (0.81,1.14)	0.88 (0.75,1.03)	0.93 (0.76,1.14)	0.90 (0.74,1.09)	0.74 (0.59,0.92)	0.81 (0.63,1.04)	0.71 (0.56,0.89)	RFM-ACE-HI	
0.96 (0.81,1.15)	0.88 (0.75,1.04)	0.94 (0.76,1.15)	0.90 (0.74,1.10)	0.74 (0.59,0.93)	0.81 (0.65,1.01)	0.71 (0.54,0.93)	1.01 (0.82,1.23)	SC-TSI

ARTICLE IN PRESS

1.16 (0.99 ,1.36)	1.07 (0.90 ,1.26)	1.13 (0.92 ,1.40)	1.09 (0.89 ,1.33)	0.89 (0.73 ,1.10)	0.98 (0.76 ,1.27)	0.85 (0.65 ,1.12)	1.21 (0.99 ,1.48)	1.21 (0.98 ,1.48)	SMI-VI							
1.03 (0.81 ,1.30)	0.94 (0.74 ,1.20)	1.00 (0.77 ,1.31)	0.97 (0.75 ,1.25)	0.79 (0.60 ,1.05)	0.87 (0.64 ,1.18)	0.76 (0.55 ,1.04)	1.07 (0.83 ,1.40)	1.07 (0.82 ,1.39)	0.89 (0.68 ,1.15)	SS						
1.10 (0.81 ,1.50)	1.01 (0.75 ,1.37)	1.07 (0.77 ,1.49)	1.03 (0.75 ,1.42)	0.85 (0.60 ,1.19)	0.93 (0.65 ,1.33)	0.81 (0.56 ,1.18)	1.15 (0.83 ,1.59)	1.14 (0.83 ,1.58)	0.95 (0.68 ,1.32)	1.07 (0.74 ,1.55)	SWA					
1.07 (0.92 ,1.23)	0.98 (0.86 ,1.11)	1.04 (0.89 ,1.22)	1.00 (0.85 ,1.18)	0.82 (0.67 ,1.01)	0.90 (0.71 ,1.14)	0.79 (0.61 ,1.01)	1.12 (0.94 ,1.33)	1.11 (0.93 ,1.33)	0.92 (0.77 ,1.10)	1.04 (0.81 ,1.33)	0.97 (0.71 ,1.32)	TAI				
1.11 (0.96 ,1.29)	1.02 (0.89 ,1.17)	1.08 (0.92 ,1.28)	1.04 (0.88 ,1.24)	0.86 (0.70 ,1.05)	0.94 (0.74 ,1.19)	0.82 (0.64 ,1.06)	1.16 (0.98 ,1.38)	1.16 (0.96 ,1.38)	0.96 (0.80 ,1.15)	1.08 (0.84 ,1.39)	1.01 (0.74 ,1.38)	1.04 (0.92 ,1.18)	TSI			
1.12 (0.93 ,1.34)	1.03 (0.87 ,1.22)	1.09 (0.88 ,1.35)	1.05 (0.86 ,1.28)	0.86 (0.68 ,1.09)	0.94 (0.73 ,1.22)	0.82 (0.63 ,1.08)	1.17 (0.95 ,1.43)	1.16 (0.94 ,1.43)	0.96 (0.78 ,1.19)	1.09 (0.83 ,1.42)	1.02 (0.73 ,1.41)	1.05 (0.87 ,1.26)	1.00 (0.83 ,1.21)	UAP		
1.07 (0.96 ,1.20)	0.98 (0.91 ,1.07)	1.04 (0.89 ,1.22)	1.01 (0.88 ,1.15)	0.83 (0.69 ,0.99)	0.91 (0.73 ,1.12)	0.79 (0.62 ,1.00)	1.12 (0.97 ,1.30)	1.11 (0.96 ,1.30)	0.92 (0.79 ,1.08)	1.04 (0.83 ,1.31)	0.97 (0.72 ,1.31)	1.00 (0.90 ,1.12)	0.96 (0.86 ,1.09)	0.96 (0.82 ,1.13)	UDFF	
1.07 (0.90 ,1.28)	0.98 (0.83 ,1.16)	1.04 (0.85 ,1.28)	1.00 (0.83 ,1.22)	0.82 (0.66 ,1.03)	0.90 (0.70 ,1.17)	0.79 (0.60 ,1.03)	1.12 (0.92 ,1.36)	1.11 (0.91 ,1.36)	0.92 (0.75 ,1.13)	1.04 (0.80 ,1.36)	0.97 (0.70 ,1.35)	1.00 (0.84 ,1.19)	0.96 (0.80 ,1.15)	0.96 (0.78 ,1.18)	1.00 (0.86 ,1.16)	USFF

Note: UDFF: Ultrasound-Derived Fat Fraction; CAP: Controlled Attenuation Parameter; HRR: Hepatorenal Ratio; AC: Attenuation Coefficient; NLV: Normalized Local Variance; SMI-VI: Superb Microvascular Imaging Vascular Index; RFM-ACE-FI: Reference Frequency Method-based Attenuation Coefficient Estimation in Fundamental Imaging; RFM-ACE-HI: Reference Frequency Method-based Attenuation Coefficient Estimation in Harmonic Imaging; AC-TAI: Attenuation Coefficient - Tissue Attenuation Imaging; SC-TSI: Scatterer Coefficient - Tissue Scatterer Imaging; SS: Sound Speed; Deep-USFF: Deep-learning based Ultrasound Fat Fraction Estimation; TAI : Tissue Attenuation Imaging; TSI: Tissue Scatterer Imaging; UAP: Ultrasound Attenuation Parameter; USFF: Ultrasound Fat Fraction Estimation; SWA: Shear Wave Attenuation; MRI-PDFF : Magnetic Resonance Imaging - Proton Density Fat Fraction.

Table 4. Relative effects and 95% Credible Intervals of all pairwise panels for SEN based on the Bayesian NMA method (Moderate steatosis/ S2).

MRI-PDFF				
1.06 (1.02 ,1.10)	USFF			
1.08 (1.01 ,1.16)	1.02 (0.94 ,1.11)	TSI		
1.08 (1.04 ,1.12)	1.03 (0.97 ,1.08)	1.00 (0.93 ,1.09)	UDFF	
1.08 (0.88 ,1.33)	1.02 (0.83 ,1.26)	1.00 (0.80 ,1.25)	1.00 (0.81 ,1.23)	RFM-ACE-HI

ARTICLE IN PRESS

1.10 (1.01 ,1.19)	1.04 (0.95 ,1.14)	1.02 (0.92 ,1.13)	1.01 (0.93 ,1.11)	1.02 (0.81 ,1.27)	UAP											
1.12 (0.98 ,1.28)	1.06 (0.93 ,1.21)	1.04 (0.90 ,1.20)	1.03 (0.90 ,1.18)	1.04 (0.81 ,1.33)	1.02 (0.87 ,1.19)	SMI-VI										
1.12 (0.99 ,1.27)	1.06 (0.93 ,1.21)	1.04 (0.91 ,1.19)	1.04 (0.91 ,1.18)	1.04 (0.82 ,1.32)	1.02 (0.88 ,1.18)	1.00 (0.84 ,1.20)	Deep-USFF									
1.13 (0.81 ,1.58)	1.07 (0.77 ,1.50)	1.05 (0.75 ,1.47)	1.05 (0.75 ,1.46)	1.05 (0.71 ,1.55)	1.03 (0.73 ,1.45)	1.01 (0.71 ,1.44)	1.01 (0.71 ,1.44)	SWA								
1.13 (1.03 ,1.25)	1.07 (0.97 ,1.19)	1.05 (0.93 ,1.18)	1.05 (0.94 ,1.16)	1.05 (0.83 ,1.32)	1.03 (0.91 ,1.17)	1.01 (0.86 ,1.19)	1.01 (0.86 ,1.18)	1.00 (0.71 ,1.41)	SC-TSI							
1.15 (1.01 ,1.31)	1.09 (0.95 ,1.24)	1.07 (0.92 ,1.23)	1.06 (0.93 ,1.21)	1.07 (0.83 ,1.36)	1.05 (0.90 ,1.22)	1.03 (0.87 ,1.21)	1.03 (0.86 ,1.22)	1.02 (0.71 ,1.45)	1.02 (0.86 ,1.19)	AC						
1.16 (1.00 ,1.34)	1.10 (0.94 ,1.28)	1.07 (0.91 ,1.26)	1.07 (0.92 ,1.24)	1.07 (0.83 ,1.38)	1.06 (0.89 ,1.25)	1.03 (0.86 ,1.24)	1.03 (0.86 ,1.25)	1.02 (0.71 ,1.47)	1.02 (0.86 ,1.22)	1.01 (0.84 ,1.21)	NLV					
1.15 (1.08 ,1.23)	1.09 (1.01 ,1.17)	1.07 (0.97 ,1.17)	1.06 (0.99 ,1.14)	1.07 (0.86 ,1.32)	1.05 (0.94 ,1.16)	1.03 (0.89 ,1.19)	1.03 (0.89 ,1.18)	1.02 (0.72 ,1.42)	1.02 (0.90 ,1.14)	1.00 (0.87 ,1.15)	0.99 (0.85 ,1.16)	CAP				
1.18 (1.01 ,1.38)	1.12 (0.96 ,1.31)	1.10 (0.93 ,1.30)	1.09 (0.94 ,1.27)	1.10 (0.85 ,1.42)	1.08 (0.90 ,1.28)	1.06 (0.86 ,1.29)	1.06 (0.87 ,1.29)	1.04 (0.72 ,1.51)	1.04 (0.87 ,1.26)	1.03 (0.84 ,1.26)	1.02 (0.83 ,1.26)	1.03 (0.88 ,1.20)	HRR			
1.24 (1.11 ,1.38)	1.17 (1.04 ,1.31)	1.14 (1.01 ,1.29)	1.14 (1.02 ,1.28)	1.14 (0.91 ,1.45)	1.12 (0.98 ,1.29)	1.10 (0.93 ,1.31)	1.10 (0.94 ,1.29)	1.09 (0.77 ,1.55)	1.09 (0.94 ,1.26)	1.07 (0.91 ,1.27)	1.07 (0.89 ,1.28)	1.07 (0.95 ,1.22)	1.04 (0.86 ,1.26)	TAI		
1.59 (1.03 ,2.45)	1.50 (0.97 ,2.32)	1.47 (0.95 ,2.28)	1.47 (0.95 ,2.26)	1.47 (0.93 ,2.33)	1.44 (0.93 ,2.25)	1.42 (0.90 ,2.23)	1.42 (0.90 ,2.22)	1.40 (0.81 ,2.42)	1.40 (0.90 ,2.19)	1.38 (0.88 ,2.17)	1.37 (0.87 ,2.16)	1.38 (0.89 ,2.14)	1.34 (0.85 ,2.12)	1.29 (0.82 ,2.01)	RFM-ACE-FI	
1.74 (1.02 ,2.96)	1.64 (0.96 ,2.80)	1.61 (0.94 ,2.75)	1.60 (0.94 ,2.73)	1.61 (0.91 ,2.85)	1.58 (0.92 ,2.71)	1.55 (0.90 ,2.68)	1.55 (0.90 ,2.67)	1.53 (0.82 ,2.87)	1.53 (0.89 ,2.63)	1.51 (0.87 ,2.60)	1.50 (0.86 ,2.60)	1.51 (0.88 ,2.58)	1.47 (0.84 ,2.55)	1.40 (0.82 ,2.42)	1.09 (0.55 ,2.17)	SS
1.55 (1.28 ,1.87)	1.46 (1.20 ,1.78)	1.43 (1.17 ,1.75)	1.43 (1.17 ,1.73)	1.43 (1.08 ,1.90)	1.41 (1.14 ,1.73)	1.38 (1.10 ,1.74)	1.38 (1.10 ,1.73)	1.36 (0.93 ,2.00)	1.36 (1.10 ,1.69)	1.34 (1.07 ,1.69)	1.33 (1.05 ,1.69)	1.34 (1.10 ,1.64)	1.31 (1.02 ,1.67)	1.25 (1.01 ,1.56)	0.97 (0.61 ,1.56)	0.89 (0.51 ,1.57)

Note: UDFF: Ultrasound-Derived Fat Fraction; CAP: Controlled Attenuation Parameter; HRR: Hepatorenal Ratio; AC: Attenuation Coefficient; NLV: Normalized Local Variance; SMI-VI: Superb Microvascular Imaging Vascular Index; RFM-ACE-FI: Reference Frequency Method-based Attenuation Coefficient Estimation in Fundamental Imaging; RFM-ACE-HI: Reference Frequency Method-based Attenuation Coefficient Estimation in Harmonic Imaging; AC-TAI: Attenuation Coefficient - Tissue Attenuation Imaging; SC-TSI: Scatterer Coefficient - Tissue Scatterer Imaging; SS: Sound Speed; Deep-USFF: Deep-learning based Ultrasound Fat Fraction Estimation; TAI: Tissue Attenuation Imaging; TSI: Tissue Scatterer Imaging; UAP: Ultrasound Attenuation Parameter; USFF :

ARTICLE IN PRESS

Ultrasound Fat Fraction Estimation; SWA: Shear Wave Attenuation; MRI-PDFF: Magnetic Resonance Imaging - Proton Density Fat Fraction.

Table 5. Relative effects and 95% CrIs of all pairwise panels for SPE based on the Bayesian NMA method (Moderate steatosis/ S2).

MRI-PDFF												
1.04 (0.81,1.34)	Deep-USFF											
1.06 (0.80,1.41)	1.02 (0.70,1.48)	RFM-ACE-HI										
1.07 (0.80,1.41)	1.02 (0.70,1.49)	1.00 (0.67,1.49)	AC-TAI									
1.07 (0.81,1.42)	1.03 (0.71,1.50)	1.01 (0.76,1.35)	1.01 (0.68,1.50)	RFM-ACE-FI								
1.09 (0.92,1.28)	1.05 (0.81,1.35)	1.03 (0.74,1.42)	1.02 (0.74,1.42)	1.01 (0.73,1.41)	TAI							
1.10 (0.93,1.30)	1.05 (0.78,1.42)	1.03 (0.74,1.43)	1.03 (0.74,1.43)	1.02 (0.74,1.42)	1.01 (0.80,1.27)	AC						
1.11 (0.84,1.47)	1.07 (0.73,1.56)	1.05 (0.71,1.56)	1.05 (0.70,1.56)	1.04 (0.70,1.54)	1.02 (0.74,1.41)	1.02 (0.73,1.41)	USFF					
1.16 (0.87,1.54)	1.11 (0.76,1.63)	1.09 (0.73,1.63)	1.09 (0.81,1.46)	1.08 (0.72,1.61)	1.06 (0.76,1.48)	1.05 (0.75,1.47)	1.04 (0.69,1.55)	SC-TSI				
1.19 (0.97,1.45)	1.14 (0.83,1.57)	1.12 (0.79,1.58)	1.12 (0.79,1.57)	1.11 (0.78,1.56)	1.09 (0.84,1.41)	1.08 (0.88,1.33)	1.07 (0.76,1.50)	1.03 (0.72,1.46)	SMI-VI			
1.21 (0.86,1.69)	1.16 (0.76,1.76)	1.14 (0.73,1.77)	1.13 (0.73,1.76)	1.12 (0.72,1.75)	1.11 (0.76,1.61)	1.10 (0.75,1.61)	1.08 (0.70,1.68)	1.04 (0.67,1.63)	1.01 (0.68,1.51)	SWA		
1.28 (0.92,1.79)	1.23 (0.81,1.86)	1.21 (0.78,1.87)	1.21 (0.78,1.86)	1.19 (0.77,1.84)	1.18 (0.81,1.70)	1.17 (0.83,1.64)	1.15 (0.75,1.77)	1.11 (0.72,1.72)	1.08 (0.75,1.57)	1.07 (0.61,1.71)	SS	
1.29 (1.16,1.43)	1.23 (0.94,1.62)	1.21 (0.90,1.63)	1.21 (0.89,1.63)	1.20 (0.89,1.62)	1.18 (0.97,1.43)	1.17 (0.96,1.43)	1.15 (0.86,1.55)	1.11 (0.82,1.51)	1.08 (0.87,1.36)	1.07 (0.75,1.52)	1.00 (0.71,1.42)	UDFF
1.30 (1.09,1.5	1.25 (0.96,1.6	1.22 (0.88,1.7	1.22 (0.88,1.7	1.21 (0.87,1.6	1.19 (1.00,1.4	1.18 (0.93,1.5	1.16 (0.84,1.6	1.12 (0.80,1.5	1.09 (0.84,1.4	1.08 (0.74,1.5	1.01 (0.70,1.4	1.01 (0.83,1.2	TSI

ARTICLE IN PRESS

ARTICLE IN PRESS

4)	2)	0)	0)	8)	2)	1)	2)	7)	2)	8)	7)	3)					
1.38 (1.08,1.77)	1.32 (0.93,1.89)	1.30 (0.89,1.90)	1.30 (0.89,1.89)	1.29 (0.88,1.87)	1.27 (0.94,1.71)	1.26 (0.93,1.70)	1.24 (0.85,1.80)	1.20 (0.82,1.75)	1.16 (0.85,1.60)	1.15 (0.75,1.75)	1.08 (0.71,1.63)	1.07 (0.83,1.39)	1.06 (0.79,1.44)	HRR			
1.43 (1.07,1.91)	1.37 (0.93,2.01)	1.34 (0.90,2.02)	1.34 (0.89,2.01)	1.33 (0.88,1.99)	1.31 (0.94,1.83)	1.30 (0.93,1.82)	1.28 (0.85,1.92)	1.23 (0.82,1.86)	1.20 (0.84,1.71)	1.18 (0.76,1.85)	1.11 (0.72,1.73)	1.11 (0.81,1.51)	1.10 (0.78,1.54)	1.03 (0.70,1.52)	UAP		
1.40 (1.22,1.62)	1.34 (1.01,1.80)	1.32 (0.96,1.81)	1.32 (0.96,1.80)	1.31 (0.95,1.79)	1.29 (1.04,1.60)	1.28 (1.02,1.59)	1.26 (0.92,1.72)	1.21 (0.88,1.67)	1.18 (0.92,1.51)	1.16 (0.80,1.68)	1.09 (0.76,1.57)	1.09 (0.92,1.29)	1.08 (0.86,1.35)	1.01 (0.78,1.32)	0.98 (0.71,1.36)	CAP	
1.42 (1.15,1.77)	1.37 (0.98,1.90)	1.34 (0.94,1.91)	1.34 (0.94,1.91)	1.33 (0.93,1.89)	1.31 (1.00,1.71)	1.30 (1.04,1.61)	1.28 (0.92,1.82)	1.23 (0.86,1.77)	1.20 (0.95,1.51)	1.18 (0.79,1.77)	1.11 (0.76,1.62)	1.11 (0.87,1.41)	1.10 (0.83,1.45)	1.03 (0.74,1.43)	1.00 (0.69,1.44)	1.02 (0.79,1.31)	NLV

Note: UDFF: Ultrasound-Derived Fat Fraction; CAP: Controlled Attenuation Parameter; HRR: Hepatorenal Ratio; AC: Attenuation Coefficient; NLV: Normalized Local Variance; SMI-VI: Superb Microvascular Imaging Vascular Index; RFM-ACE-FI: Reference Frequency Method-based Attenuation Coefficient Estimation in Fundamental Imaging; RFM-ACE-HI: Reference Frequency Method-based Attenuation Coefficient Estimation in Harmonic Imaging; AC-TAI: Attenuation Coefficient - Tissue Attenuation Imaging; SC-TSI: Scatterer Coefficient - Tissue Scatterer Imaging; SS: Sound Speed; Deep-USFF: Deep-learning based Ultrasound Fat Fraction Estimation; TAI: Tissue Attenuation Imaging; TSI: Tissue Scatterer Imaging; UAP: Ultrasound Attenuation Parameter; USFF: Ultrasound Fat Fraction Estimation; SWA: Shear Wave Attenuation; MRI-PDFF : Magnetic Resonance Imaging - Proton Density Fat Fraction.

Table 6. Relative effects and 95% CrIs of all pairwise panels for SEN based on the Bayesian NMA method (Severe steatosis/ S3).

MRI-PDFF								
1.00 (0.87,1.14)	TAI							
1.01 (0.84,1.22)	1.01 (0.84,1.22)	Deep-USFF						
1.02 (0.89,1.17)	1.02 (0.85,1.24)	1.01 (0.80,1.27)	SS					
1.03 (0.88,1.20)	1.03 (0.88,1.20)	1.02 (0.84,1.24)	1.01 (0.82,1.24)	TSI				
1.04 (0.92,1.17)	1.04 (0.87,1.24)	1.03 (0.82,1.28)	1.01 (0.85,1.21)	1.01 (0.83,1.22)	SMI-VI			
1.05 (0.95,1.16)	1.05 (0.88,1.24)	1.04 (0.84,1.28)	1.02 (0.89,1.17)	1.01 (0.84,1.22)	1.01 (0.88,1.16)	AC		
1.06 (0.88,1.28)	1.06 (0.84,1.33)	1.05 (0.81,1.36)	1.04 (0.83,1.30)	1.03 (0.81,1.31)	1.02 (0.82,1.27)	1.01 (0.82,1.25)	USFF	
1.10 (0.95,1.28)	1.10 (0.90,1.34)	1.09 (0.86,1.38)	1.08 (0.89,1.30)	1.07 (0.86,1.32)	1.06 (0.90,1.25)	1.05 (0.90,1.24)	1.04 (0.82,1.32)	NLV

ARTICLE IN PRESS

1.11 (0.94,1 .31)	1.11 (0.90,1 .37)	1.10 (0.86,1 .41)	1.09 (0.88,1 .34)	1.08 (0.86,1 .35)	1.07 (0.87,1 .31)	1.06 (0.88,1 .29)	1.05 (0.82,1 .34)	1.01 (0.81,1 .26)	HRR				
1.12 (0.91,1 .38)	1.12 (0.87,1 .44)	1.11 (0.84,1 .47)	1.10 (0.85,1 .41)	1.09 (0.84,1 .41)	1.08 (0.85,1 .38)	1.07 (0.85,1 .35)	1.06 (0.80,1 .40)	1.02 (0.79,1 .32)	1.01 (0.77,1 .32)	SWA			
1.11 (0.96,1 .29)	1.11 (0.91,1 .36)	1.10 (0.87,1 .40)	1.09 (0.89,1 .33)	1.08 (0.87,1 .33)	1.07 (0.89,1 .30)	1.06 (0.89,1 .27)	1.05 (0.83,1 .33)	1.01 (0.82,1 .24)	1.00 (0.80,1 .25)	0.99 (0.77,1 .28)	UAP		
1.19 (1.10,1 .27)	1.19 (1.02,1 .38)	1.17 (0.96,1 .43)	1.16 (1.00,1 .35)	1.15 (0.97,1 .36)	1.14 (1.00,1 .31)	1.13 (1.00,1 .28)	1.12 (0.92,1 .36)	1.08 (0.92,1 .26)	1.07 (0.90,1 .27)	1.06 (0.85,1 .32)	1.07 (0.90,1 .26)	UDFF	
1.23 (1.09,1 .39)	1.23 (1.03,1 .47)	1.22 (0.97,1 .52)	1.20 (1.00,1 .44)	1.19 (0.98,1 .45)	1.18 (1.00,1 .40)	1.17 (1.00,1 .37)	1.16 (0.93,1 .44)	1.11 (0.92,1 .35)	1.11 (0.92,1 .34)	1.10 (0.86,1 .40)	1.10 (0.91,1 .34)	1.04 (0.91,1 .18)	CAP

Note: UDFF: Ultrasound-Derived Fat Fraction; CAP: Controlled Attenuation Parameter; HRR: Hepatorenal Ratio; AC: Attenuation Coefficient; NLV: Normalized Local Variance; SMI-VI: Superb Microvascular Imaging Vascular Index; SS: Sound Speed; Deep-USFF: Deep-learning based Ultrasound Fat Fraction Estimation; TAI: Tissue Attenuation Imaging; TSI: Tissue Scatterer Imaging; UAP: Ultrasound Attenuation Parameter; USFF: Ultrasound Fat Fraction Estimation; SWA: Shear Wave Attenuation; MRI-PDFF : Magnetic Resonance Imaging - Proton Density Fat Fraction.

Table 7. Relative effects and 95% CrIs of all pairwise panels for SPE based on the Bayesian NMA method (Severe steatosis/ S3).

MRI-PDFF											
1.03 (0.65,1. 64)	Deep-USFF										
1.10 (0.93,1. 31)	1.07 (0.65,1. 75)	UDFF									
1.07 (0.64,1. 78)	1.03 (0.52,2. 06)	0.97 (0.56,1. 66)	SWA								
1.16 (0.81,1. 66)	1.13 (0.71,1. 79)	1.05 (0.71,1. 57)	1.09 (0.58,2. 03)	TAI							
1.17 (0.71,1. 94)	1.14 (0.58,2. 25)	1.06 (0.63,1. 81)	1.10 (0.54,2. 25)	1.01 (0.55,1. 87)	USFF						
1.32 (0.80,2. 18)	1.28 (0.64,2. 53)	1.19 (0.70,2. 04)	1.23 (0.60,2. 53)	1.13 (0.61,2. 10)	1.12 (0.55,2. 29)	UAP					
1.33 (0.87,2. 04)	1.29 (0.69,2. 42)	1.21 (0.78,1. 86)	1.25 (0.64,2. 43)	1.14 (0.66,2. 00)	1.13 (0.59,2. 19)	1.01 (0.52,1. 96)	HRR				
1.35 (0.95,1. 90)	1.30 (0.73,2. 32)	1.22 (0.83,1. 80)	1.26 (0.68,2. 34)	1.16 (0.70,1. 91)	1.15 (0.62,2. 11)	1.02 (0.55,1. 89)	1.01 (0.58,1. 75)	SMI-VI			
1.33 (1.07,1. 67)	1.29 (0.77,2. 16)	1.21 (0.92,1. 58)	1.25 (0.71,2. 18)	1.15 (0.75,1. 75)	1.14 (0.66,1. 97)	1.01 (0.58,1. 76)	1.00 (0.64,1. 56)	0.99 (0.66,1. 50)	CAP		
1.42 (0.99,2. 05)	1.38 (0.86,2. 21)	1.29 (0.86,1. 93)	1.33 (0.71,2. 50)	1.23 (0.85,1. 77)	1.21 (0.65,2. 25)	1.08 (0.58,2. 01)	1.07 (0.61,1. 87)	1.06 (0.64,1. 75)	1.07 (0.70,1. 64)	TSI	
1.60 (1.18,2.	1.55 (0.89,2.	1.45 (1.02,2.	1.50 (0.83,2.	1.38 (0.86,2.	1.36 (0.76,2.	1.21 (0.67,2.	1.20 (0.71,2.	1.19 (0.84,1.	1.20 (0.82,1.	1.12 (0.70,1.	AC

ARTICLE IN PRESS

17)	69)	06)	71)	20)	45)	19)	03)	68)	75)	80)			
2.10 (1.46,3.02)	2.03 (1.13,3.66)	1.90 (1.27,2.85)	1.97 (1.05,3.68)	1.80 (1.08,3.01)	1.79 (0.96,3.32)	1.59 (0.85,2.97)	1.58 (0.90,2.76)	1.56 (1.06,2.28)	1.57 (1.03,2.41)	1.47 (0.88,2.46)	1.31 (0.91,1.90)	NLV	
2.42 (1.43,4.11)	2.34 (1.16,4.72)	2.19 (1.26,3.83)	2.26 (1.08,4.73)	2.08 (1.10,3.94)	2.06 (0.99,4.27)	1.84 (0.88,3.82)	1.82 (0.92,3.59)	1.80 (0.98,3.29)	1.81 (1.02,3.22)	1.70 (0.89,3.23)	1.51 (0.87,2.62)	1.15 (0.62,2.13)	SS

Note: UDFF: Ultrasound-Derived Fat Fraction; CAP: Controlled Attenuation Parameter; HRR: Hepatorenal Ratio; AC: Attenuation Coefficient; NLV: Normalized Local Variance; SMI-VI: Superb Microvascular Imaging Vascular Index; SS: Sound Speed; Deep-USFF: Deep-learning based Ultrasound Fat Fraction Estimation; TAI: Tissue Attenuation Imaging; TSI : Tissue Scatterer Imaging; UAP: Ultrasound Attenuation Parameter; USFF: Ultrasound Fat Fraction Estimation; SWA: Shear Wave Attenuation; MRI-PDFF: Magnetic Resonance Imaging - Proton Density Fat Fraction.

Figure legends

Figure 1. Study selection flowchart.

Figure 2. The quality assessment of the included studies via the QUADAS-2 tool. (A) RoB and applicability concerns graph. (B) Summary of quality assessment.

Figure 3. Network diagram. (A) A head-to-head comparison of the diagnostic SEN of different ultrasound tools for mild steatosis, using MRI-PDFF as the gold standard (S1_sen). (B) A head-to-head comparison of the diagnostic SPE of different ultrasound tools for mild steatosis, using MRI-PDFF as the gold standard (S1_spe). (C) A head-to-head comparison of the diagnostic SEN of different ultrasound tools for moderate steatosis, using MRI-PDFF as the gold standard (S2_sen). (D) A head-to-head comparison of the diagnostic SPE of different ultrasound tools for moderate steatosis, using MRI-PDFF as the gold standard (S2_spe). (E) A head-to-head comparison of the diagnostic SEN of different ultrasound tools for severe steatosis, using MRI-PDFF as the gold standard (S3_sen). (F) A head-to-head comparison of the diagnostic SPE of different ultrasound tools for severe steatosis, using MRI-PDFF as the gold standard (S3_spe).

Figure 4. Cumulative probability plot. (A) The SEN of each tool in diagnosing mild steatosis (S1_sen). (B) The SPE of each tool in diagnosing mild steatosis (S1_spe). (C) The SEN of each tool in diagnosing moderate steatosis (S2_sen). (D) The SPE of each tool in diagnosing moderate

ARTICLE IN PRESS

steatosis (S2_spe). (E) The SEN of each tool in diagnosing severe steatosis (S3_sen). (F) The SPE of each tool in diagnosing severe steatosis (S3_spe).

Figure 5. Comparative adjusted funnel plot for publication bias (Each point in the picture is the result of a pairwise comparison between the tools). (A) Publication bias across the 16 studies evaluating the diagnostic SEN of ultrasound tools for mild steatosis (S1_sen). (B) Publication bias across the 16 studies evaluating the diagnostic SPE of ultrasound tools for mild steatosis (S1_spe). (C) Publication bias across the 15 studies evaluating the diagnostic SEN of ultrasound tools for moderate steatosis (S2_sen). (D) Publication bias across the 15 studies evaluating the diagnostic SPE of ultrasound tools for moderate steatosis (S2_spe). (E) Publication bias across the 12 studies evaluating the diagnostic SEN of ultrasound tools for severe steatosis (S3_sen). (F) Publication bias across the 12 studies evaluating the diagnostic SPE of ultrasound tools for severe steatosis (S3_spe).

ARTICLE IN PRESS

Supplementary Materials

Supplementary Table S1. Literature search strategy.

Supplementary Table S2. Outcome information.

Supplementary Table S3. Detailed information on index tests.

Supplementary Table S4. Local inconsistency test.

Supplementary Table S5. ANOVA model results.

Supplementary Table S6. Surface under the cumulative ranking curve.