European Radiology (2019) 29:5180–5189
https://doi.org/10.1007/s00330-019-06071-5

MAGNETIC RESONANCE

Diagnostic accuracy of hepatic proton density fat fraction measured by magnetic resonance imaging for the evaluation of liver steatosis with histology as reference standard: a meta-analysis

Yali Qu[1] · Mou Li[1] · Gavin Hamilton[2] · Yingzhen N. Zhang[2] · Bin Song[1]

Received: 20 July 2018 / Revised: 24 December 2018 / Accepted: 4 February 2019 / Published online: 15 March 2019

Abstract

Objectives The aim of this meta-analysis was to evaluate the diagnostic accuracy of hepatic magnetic resonance imaging-proton density fat fraction (MRI-PDFF) for the assessment of liver steatosis (LS) with histology as reference standard.

Methods A systematic literature search was performed to identify pertinent studies. Quality analyses were conducted by Quality Assessment of Diagnostic Accuracy Studies-2. Diagnostic data were extracted and inconsistency index was calculated for LS≥G1, LS≥G2, and LS=G3, respectively. The area under summary receiver operating characteristic curve (AUC) served as the indicator of diagnostic accuracy. The pooled sensitivity and specificity were calculated if threshold effect was absent.

Results Thirteen studies containing 1100 subjects were included. There was significant threshold effect for LS≥G1. The AUCs for LS≥G1, LS≥G2, and LS=G3 were 0.98 (95% confidence interval (CI) 0.76, 1.00), 0.91 (95% CI 0.89, 0.94), and 0.92 (95% CI 0.89, 0.94), respectively. The pooled sensitivities for LS≥G2 and LS=G3 were 0.83 (95% CI 0.75, 0.88) and 0.79 (95% CI 0.63, 0.90), respectively; the pooled specificities for LS≥G2 and LS=G3 were 0.89 (95% CI 0.84, 0.92) and 0.89 (95% CI 0.84, 0.92), respectively.

Conclusions MRI-PDFF has high diagnostic accuracy at detecting and grading LS with histology as reference standard, suggesting that MRI-PDFF is able to provide an accurate quantification of LS in clinical trials and patient care.

Key Point

• *MRI-PDFF is able to provide an accurate quantification of LS in clinical trials and patient care.*

Keywords Fatty liver · Magnetic resonance imaging · Area under curve · Meta-analysis

Abbreviations

AUC	Area under summary receiver operating characteristic curve
CI	Confidence interval
CSE-MRI	Chemical shift–encoded magnetic resonance imaging
FN	False negative
FP	False positive
LS	Liver steatosis
MeSH	Medical Subject Headings
MRS	Magnetic resonance spectroscopy
NAFLD	Nonalcoholic fatty liver disease
NASH	Nonalcoholic steatohepatitis
PDFF	Proton density fat fraction
SEN	Sensitivity
SPE	Specificity
SROC	Summary receiver operating characteristic
TN	True negative
TP	True positive

Yali Qu and Mou Li contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00330-019-06071-5) contains supplementary material, which is available to authorized users.

✉ Bin Song
binsong65@yahoo.com

[1] Department of Radiology, West China Hospital of Sichuan University, No. 37 Guoxue Xiang, Chengdu 610041, Sichuan, China

[2] Liver Imaging Group, Department of Radiology, University of California at San Diego, La Jolla, CA, USA

Introduction

Liver steatosis (LS) is a very common entity characterized by the accumulation of lipid in hepatocytes. LS can be caused by

excessive alcohol use, viral hepatitis, cystic fibrosis, etc. but the most common contributor is nonalcoholic fatty liver disease (NAFLD) [1]. NAFLD represents a spectrum of liver diseases ranging from simple steatosis to nonalcoholic steatohepatitis (NASH) [2, 3]. Early diagnosis and severity grading of LS are important as timely intervention can prevent it from progressing to more advanced stages [2].

Histopathological visualization of hepatocellular fat droplets remains the gold standard for the assessment of LS [4–6]. However, biopsy is invasive, semi-quantitative, prone to sampling variability, and observer dependent [5–8]. These limitations hamper its widespread use for diagnosis and longitudinal monitoring.

Proton density fat fraction (PDFF) is a magnetic resonance (MR)–based biomarker of hepatic steatosis and is defined to be the proportion of MR-visible fat protons to the sum of MR-visible fat and water protons [9, 10]. Hepatic PDFF is at most minimally dependent on scanner platform, field strength, and appropriately chosen scanning parameters [1]. To estimate hepatic PDFF by advanced chemical shift–encoded MR imaging (CSE-MRI) or ^{1}H magnetic resonance spectroscopy (MRS), important confounding factors must be addressed, including T1 bias, T2 or T2* decay, spectral complexity of fat, and, for complex-based imaging sequences, also noise bias and eddy currents [1, 11]. MRS-PDFF has been the accepted noninvasive reference standard for hepatic MRI-PDFF quantification [12–17]. However, MRS only measures the hepatic fat fraction within the preselected voxel, which is a source for potential bias as the distribution of fat in the liver is heterogeneous [18, 19]. Advanced CSE-MRI can generate PDFF maps of the entire liver, which is helpful to assess the fat distribution in the liver and to co-localize the regions of interest across time points.

Multiple previous studies have focused on the accuracy of MRI-PDFF for the assessment of LS with histology [9, 10, 20–30] or MRS-PDFF [12–17, 31, 32] as reference standard. A previous meta-analysis has shown excellent linearity and bias between MRI-PDFF and MRS-PDFF [33], but comprehensive analysis of the existing data with histology as reference standard is still limited. Therefore, the aim of this meta-analysis is to evaluate the accuracy of MRI-PDFF for the evaluation of LS with histology as the reference standard.

Material and methods

Literature search

Two investigators independently performed a systematic literature search in PubMed, Embase, and Web of Science to identify relevant articles (last search update May 2018). The following terms including Medical Subject Headings (MeSH) terms and free-text words were used: "fatty liver or liver steatosis or liver steatohepatitis or nonalcoholic fatty liver disease or NAFLD or nonalcoholic steatohepatitis or NASH" and "magnetic resonance imaging or MR imaging or MRI or magnetic resonance spectroscopy or MR spectroscopy or MRS" and "fat quantification or multi echo or multiple echo or proton density fat fraction or pdff or T1 independent or T2* corrected or multipeak spectral or multipeak spectrum." The search was limited to studies published in English.

Inclusion and exclusion criteria

Inclusion criteria were as follows: (a) MRI-PDFF was performed for liver fat quantification; (b) all subjects had undergone hepatic histological analysis as the reference standard; (c) field strength of MR techniques was 1.5 and/or 3.0 T; (d) sufficient data were available for the calculation of true-positive (TP), false-positive (FP), false-negative (FN), and true-negative (TN) values; (e) 10 human individuals were evaluated at least. Authors from studies without sufficient published data were contacted for the missing data. Animal or ex vivo studies, duplicate publication, secondary analysis of previously published data, review articles, abstracts, case reports, comments, and letters were excluded.

Quality assessment and data extraction

The methodological quality of the included studies was assessed according to Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool [34]. For each study, we extracted and recorded data using a standardized data extraction form, including the following: author, publication year, study region, study design, patient characteristics, blinding procedure, the utilization of Nonalcoholic Steatohepatitis Clinical Research Network (NASH CRN) scoring system for histological grading (G0, < 5% hepatocytes; G1, 5–33% hepatocytes, G2: > 33–66% hepatocytes; G3, > 66% hepatocytes [35]), and imaging parameters of MR. Meanwhile, values for TP, FP, FN, TN, sensitivity (SEN), specificity (SPE), and diagnostic threshold of MR were also recorded.

Statistical analysis

The inconsistency index (*I*-squared, I^2) was used to estimate the heterogeneity among the enrolled individual studies. $I^2 = 100\% \times (Q - \mathrm{df})\ Q$, where Q is the statistic of the chi-square test and df is the degrees of freedom [36]. I^2 indicates the percentage of total variation across studies that is attributable to heterogeneity rather than to chance [37]. $I^2 > 50\%$ indicated notable heterogeneity [36, 38]. Threshold effect means that different threshold values used by different studies to define a positive or negative test result contribute to the heterogeneity. A strong positive correlation between the logit of SEN and

the logit of (1 − SPE) with $p < 0.05$ indicated the presence of significant threshold effect, which was carried out by Meta-DiSc (version 1.4) [39]. A bivariate mixed-effect binary regression model was used to summarize the diagnostic performance [37]. Bivariate model takes into account the heterogeneity beyond chance and is commonly used when notable heterogeneity is present [40, 41]. Summary receiver operating characteristic (SROC) curve analyses were conducted, and the area under SROC curve (AUC) was regarded as the indicator of the diagnostic accuracy. SROC curve analysis is a summary method that accounts for threshold differences among original studies [42, 43]. Pooled SENs and SPEs were calculated when the threshold effect was absent [41], which were presented alongside the SENs and SPEs from each included study using forest plots. Sensitivity analyses were performed to evaluate the stability and strength of results. Deeks' funnel plot asymmetry test was used to evaluate the publication bias which was considered as present in the absence of zero slope coefficient ($p < 0.05$) [44]. All of the above analyses except for threshold effect assessment were performed using the MIDAS command of Stata (version 12.0).

Results

Search results

The search yielded 1609 literature citations in all. Of these, the full texts of 221 articles were read. Finally, 13 eligible articles estimating the accuracy of MRI-PDFF for the assessment of LS with histology as reference standard were included. Our efforts to obtain the patient-level data were not successful. The procedure of the systematic search is outlined in Fig. 1.

Study characteristics and quality assessment

Of the 13 included studies, 11 employed NASH CRN histological scoring system [9, 10, 20–25, 27–29] and 2 defined LS as a ≥ 5% fatty hepatocyte without steatosis grading [26, 30]. Therefore, these 2 studies were included in the analysis of detecting LS≥G1, but excluded in the analyses of detecting LS≥G2 and LS=G3. As a result, there were 8 studies (N = 728) for LS≥G1 [10, 21, 24–27, 29, 30], 9 studies (N = 719) for LS≥G2 [9, 10, 20–23, 25, 27, 28], and 6 studies (N = 553) for LS=G3 [10, 20–23, 27]. Patient characteristics, methodology, imaging protocol, and diagnostic results of included studies are listed in Tables 1, 2, and 3. The thresholds of MRI-PDFF for detecting or grading LS varied among individual studies, 3.42–6.90% for LS≥G1, 10.00–17.50% for LS≥G2, and 16.37–23.50% for LS=G3 (Table 3). The quality of the included studies was good (Fig. 2).

Assessment of heterogeneity

The heterogeneity tests for LS≥G1, LS≥G2, and LS=G3 showed I^2 = 87%, 46%, and 69%, respectively, indicating the presence of notable heterogeneity in LS≥G1 and LS=G3. For LS≥G1, the Spearman correlation coefficient between the logit of SEN and the logit of (1 − SPE) was 0.728 (p = 0.037), indicating the presence of threshold effect. There was no significant threshold effect among individual studies for LS≥G2 or LS=G3 ($p > 0.2$). The number of eligible studies for LS=G3 was too limited to perform meta-regression and subgroup analysis for exploring the potential contributors to the heterogeneity.

Quantitative synthesis

The AUCs for LS≥G1, LS≥G2, and LS=G3 were 0.98 (95% confidence interval (CI) 0.76, 1.00), 0.91 (95% CI 0.89, 0.94), and 0.92 (95% CI 0.89, 0.94), respectively. The pooled SENs for LS≥G2 and LS=G3 were 0.83 (95% CI 0.75, 0.88) and 0.79 (95% CI 0.63, 0.90), respectively; the pooled SPEs for LS≥G2 and LS=G3 were 0.89 (95% CI 0.84, 0.92) and 0.89 (95% CI 0.84, 0.92), respectively. The pooled SEN and SPE were not calculated for LS≥G1 since the presence of significant threshold effect. The SROC curves and forest plots of SEN and SPE are presented in Fig. 3. The results of sensitivity analysis indicated that our results were reliable, and the details are further described in Appendix in the ESM.

Publication bias

The p values of Deeks' funnel plot asymmetry tests were 0.49 for LS≥G1, 0.61 for LS≥G2, and 0.45 for LS=G3, respectively, which demonstrated no evidence of notable publication bias.

Discussion

In this meta-analysis including 13 eligible studies, our results showed that MRI-PDFF has high accuracy at diagnosing and grading LS with histology as reference standard, suggesting that MRI-PDFF is able to provide an accurate quantification of LS in clinical trials and patient care.

Multiple single- and multi-center studies have assessed the diagnostic accuracy of MRI-PDFF using histology as reference standard [9, 10, 20–30]. However, there has been no meta-analysis of these studies to date. Study design, population, MR parameters, and results varied among those prior studies. Our meta-analysis summarized the results of those individual studies with a systematic and statistical method, generating the pooled estimates that provide comprehensive evidence about the accuracy of MRI-PDFF on LS assessment

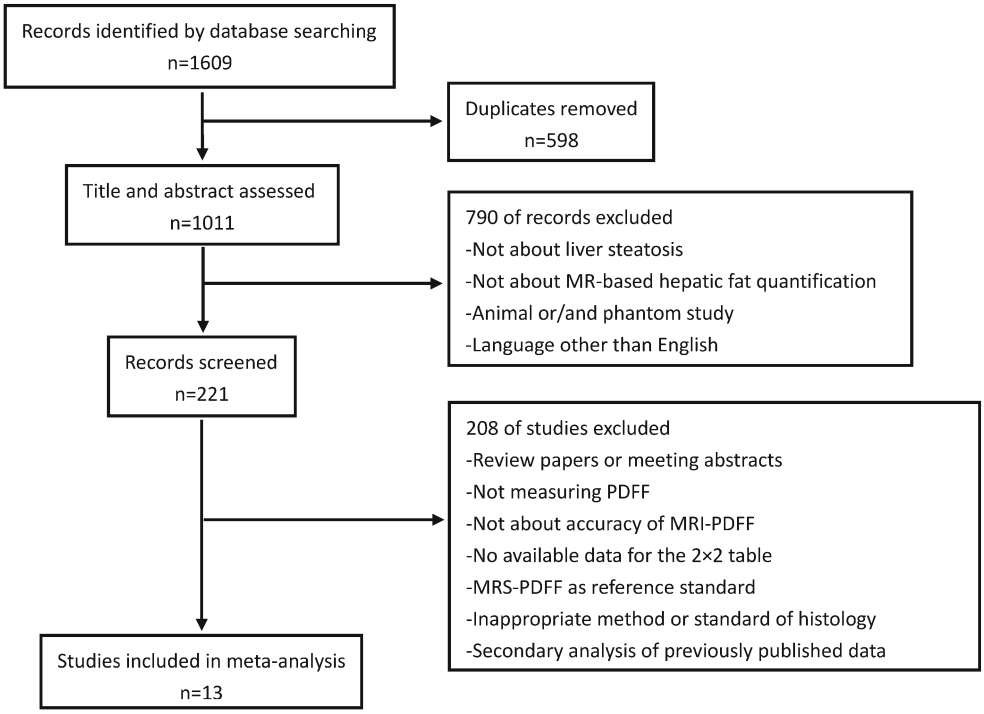

Fig. 1 Flow diagram shows the process of study selection with reasons for excluding certain studies

for future clinical trials and patient care. Bohte et al [45] performed a meta-analysis of the accuracy of signal fat-fraction measured by dual-echo MRI for the assessment of LS with histology as reference standard. Signal fat-fraction refers to the signal attributed to hepatic fat, which is a confounded parameter and is not a standardized biomarker for the quantification of LS [1]. Multiple original studies illustrating the correlation between MRI-PDFF and histological steatosis showed high correlation coefficients [46–49], which is consistent with our findings.

In our study, we noticed that the thresholds of MRI-PDFF for detecting or grading LS were not uniform among original studies. The threshold for detecting LS is more important than that for grading LS in clinical trials and patient care as it is vital for discriminating LS from nonsteatotic liver. A plausible reason for the heterogeneity in the thresholds of MRI-PDFF

Table 1 Study and subjects characteristics of included studies

Study	Year	Region	Study design	Sample size (m/f)	Mean age (years)	Patient spectrum (*N*)	BMI (kg/m^2) (mean ± SD)
Joe et al	2012	Korea	Retro	49 (32/17)	31.7	Potential liver donors (49)	NA
Kühn et al	2012	Germany	Pros	95 (50/45)	57.3	Elevated liver enzyme levels (51); suspected malignant liver tumors (44)	NA
Idilman et al	2013	Turkey	Retro	70 (40/30)	44.7	NAFLD (70)	29.9 ± 4.3
Tang et al	2013	USA	Pros	77 (61/16)	14.0[§]	NAFLD (77)	33.2 ± 6.0 (adults) 2.3 ± 0.4 (children)*
Chiang et al	2014	Taiwan	Pros	63 (29/34)	30.0	Living liver donors (63)	23.0 ± 4.1
Paparo et al	2015	Italy	Pros	77 (43/34)	51.3	Chronic C hepatitis (77)	22.39 ± 2.27
Schwimmer et al	2015	USA	Pros	174 (118/56)	14.0	NAFLD (150); normal (24)	33.3 ± NA
Tang et al	2015	USA	Pros	89 (38/51)	51.0	NAFLD or suspected NAFLD (89)	30.6 ± 5.0
Idilman et al	2016	Turkey	Retro	19 (15/4)	41.7	NAFLD (19)	27.5 ± 3.3
Middleton et al	2017	USA	Pros	113(43/70)	51.0	NASH (113)	33.6 ± 5.2
Paige et al	2017	USA	Pros	60 (30/30)	50.0	NAFLD (60)	32.6 ± 6.9
Park et al	2017	USA	Pros	104 (45/59)	50.8	Suspected NAFLD (104)	30.4 ± 5.2
Middleton et al	2018	USA	Pros	110 (78/32)	13.0	NAFLD (110)	32.0 ± 6.0

m/f, male-female ratio; *Retro*, retrospective; *Pros*, prospective; *NA*, data unavailable; *BMI*, body mass index

[§] Median age. *BMI *z* score

Table 2 Methodological and imaging protocol characteristics of included studies

Study	B	HG	Manufacturer	FS (T)	Dimension	FA	TR (ms)	NOE	RM
Joe 2012	Y	N	GE	1.5	3D	5°	13.7	6	H
Kühn 2012	Y	Y	Siemens	1.5	3D	10°	11.0	3	C
Idilman 2013	Y	Y	GE	1.5	3D	5°	12.9	6	C
Tang 2013	Y	Y	Siemens (1.5 T)/GE (3.0 T)	1.5/3.0	2D	10°	120–270	6	M
Chiang 2014	NA	N	GE	1.5	3D	5°	13.7	6	C
Paparo 2015	Y	Y	GE	1.5	2D	20°	120–270	16	NA
Schwimmer 2015	Y	Y	GE	3.0	2D	NA	≥ 150	6	M
Tang 2015	Y	Y	GE	3.0	2D	10°	120–270	6	M
Idilman 2016	Y	Y	GE	1.5	3D	5°	12.9	6	C
Middleton 2017	Y	Y	GE/Siemens	1.5/3.0	2D	10°	≥ 120	6	M
Paige 2017	Y	Y	GE	3.0	NA	NA	NA	6	M
Park 2017	Y	Y	GE	3.0	2D	10°	120–270	6	M
Middleton 2018	Y	Y	GE/Siemens/Phillips	1.5/3.0	2D	10°	≥ 120	6	M

B, blind; *HG*, histological grading; *FS*, field strength; *FA*, flip angle; *TR*, repetition time; *NOE*, no. of echoes; *RM*, reconstruction method; *Y*, yes; *N*, no; *NA*, data unavailable; *H*, hybrid; *C*, complex; *M*, magnitude

Table 3 Diagnostic results of MRI-PDFF for detecting and grading LS with histology as reference standard

Study	TP	FP	FN	TN	SEN	SPE	Threshold value of MRI-PDFF (%)
LS≥G1: differentiating LS from normal							
Joe 2012	16	3	0	30	1.00	0.91	5.00
Kühn 2012	44	0	7	46	0.86	1.00	5.10
Tang 2013	70	0	2	5	0.97	1.00	6.40
Chiang 2014	15	11	0	37	1.00	0.77	3.42
Paparo 2015	27	1	4	45	0.87	0.98	6.87
Schwimmer 2015	143	4	7	20	0.95	0.83	3.50
Tang 2015	70	0	13	6	0.84	1.00	6.90
Park 2017	68	0	3	7	0.96	1.00	3.71
LS≥G2: differentiating moderate or severe LS from mild or no LS							
Idilman 2013	41	4	3	22	0.93	0.85	15.03
Tang 2013	28	3	18	28	0.61	0.90	17.40
Paparo 2015	7	8	1	61	0.88	0.88	11.08
Tang 2015	34	4	10	41	0.77	0.91	16.40
Idilman 2016	12	2	0	5	1.00	0.71	10.00
Middleton 2017	62	3	13	35	0.83	0.90	16.30
Paige 2017	28	1	5	26	0.85	0.96	13.45
Park 2017	24	8	6	40	0.80	0.83	13.03
Middleton 2018	67	2	14	17	0.74	0.90	17.50
LS=G3: detection of severe LS							
Tang 2013	13	5	6	53	0.68	0.91	22.10
Tang 2015	10	5	4	70	0.71	0.93	23.50
Middleton 2017	26	8	5	74	0.84	0.90	21.70
Paige 2017	17	8	0	35	1.00	0.81	16.83
Park 2017	9	11	2	56	0.82	0.84	16.37
Middleton 2018	36	5	24	45	0.60	0.90	23.30

TP, true positive; *FP*, false positive; *FN*, false negative; *TN*, true negative; *SEN*, sensitivity; *SPE*, specificity; *NA*, data unavailable

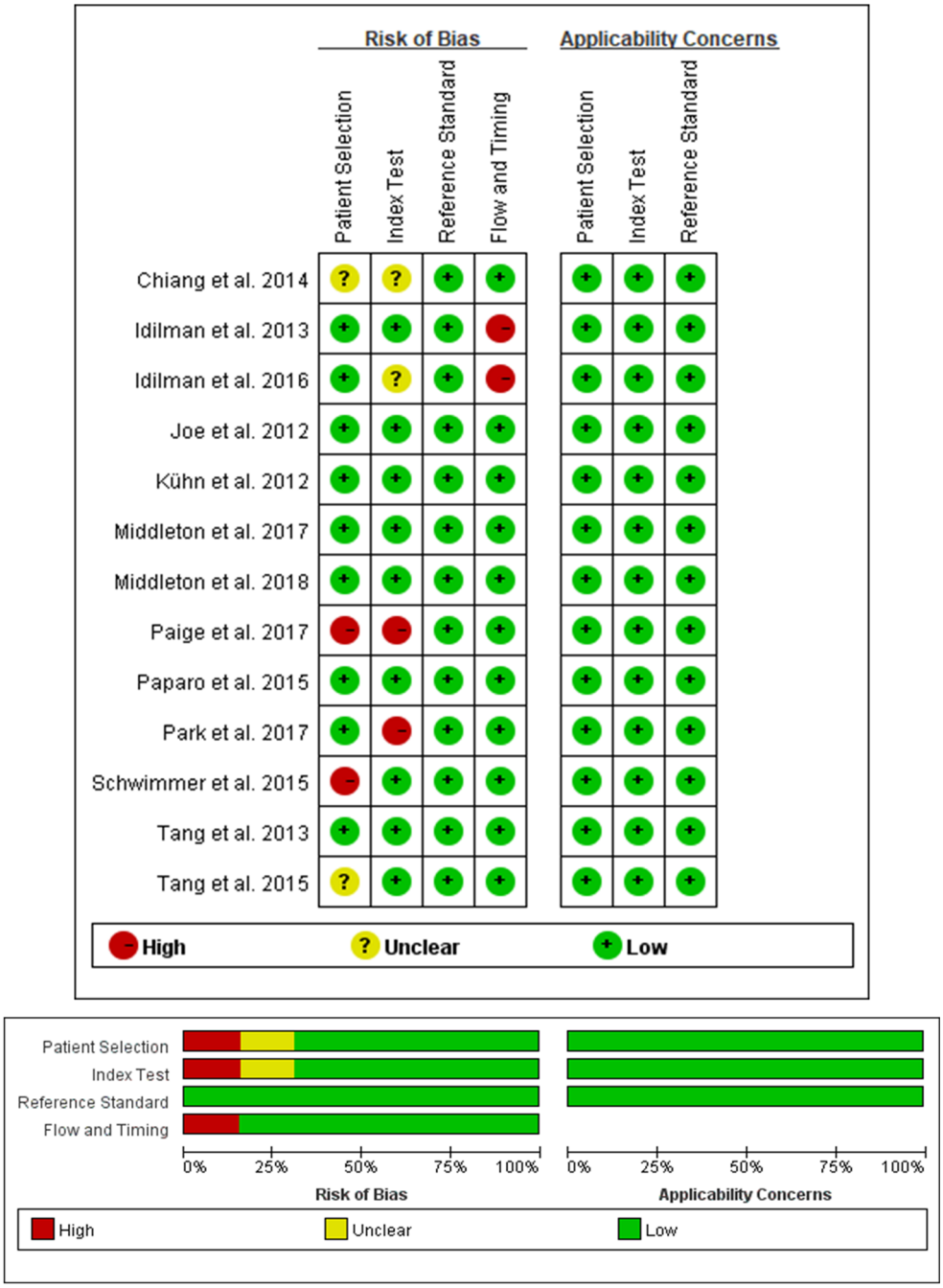

Fig. 2 Summary of methodological quality assessment for individual studies and the proportion of studies with low, high, or unclear risk of bias for each item according to QUADAS-2 tool. The results showed the quality of the included studies was good

among individual studies may be that standardized thresholds of MRI-PDFF have not been established to date. Recently, two randomized controlled trails and one study regarding MRI-PDFF as reference standard defined LS by ≥ 5% on MRI-PDFF [50–52]. Further studies are needed to establish standardized MRI-PDFF thresholds for the assessment of LS, especially the one for detecting LS.

Histologic interpretation has been used as reference standard for the assessment of LS in many clinical trials and patient care. However, histology is a suboptimal standard for multiple reasons. First, histology and MRI measure different entities for hepatic fat quantification. Histology-based fat fraction is a visual estimate of the percentage of hepatocytes containing fat droplets, while MRI-PDFF measures the proportion of MR-visible fat protons to the sum of MR-visible fat and water protons [10]. It explains why hepatic MRI-PDFF percentages are almost always less than half of histological steatosis percentages, and why hepatic MRI-PDFF basically does not exceed 50% in practice [23]. The fat droplets in a hepatocyte do not generally exceed 50%; if all examined hepatocytes are filled with 50% fat droplets on hematoxylin and eosin, the steatosis percentage would be 100% determined by

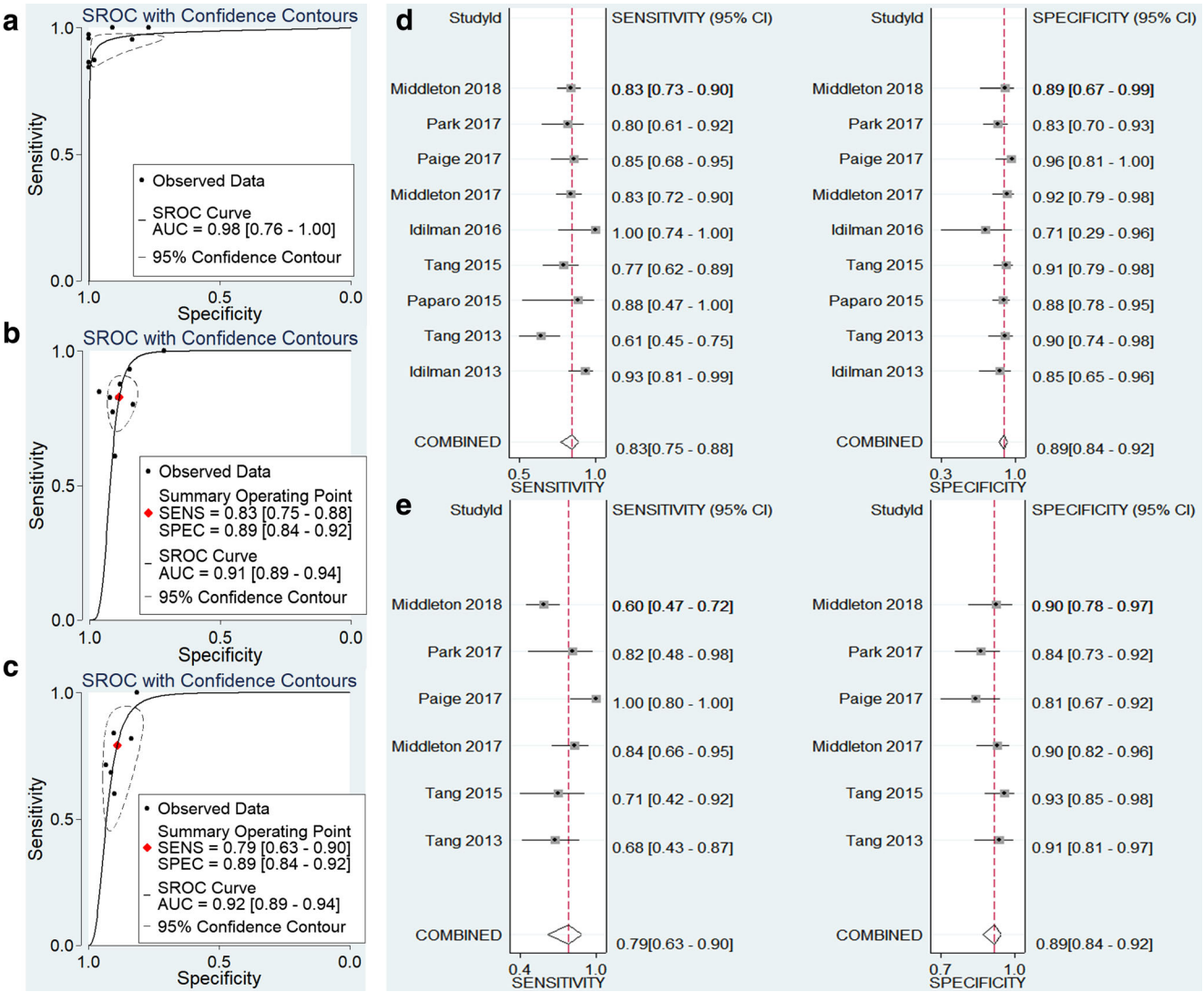

Fig. 3 SROC curves for LS≥G1 (**a**), LS≥G2 (**b**), and LS=G3 (**c**), and forest plots of sensitivity and specificity for LS≥G2 (**d**) and LS=G3 (**e**). The results show MRI-PDFF has high diagnostic accuracy for detecting and grading LS with histology as reference standard. Since the presence of significant threshold effect for LS≥G1, pooled sensitivities and specificities were calculated only for LS≥G2 (**d**) and LS=G3 (**e**)

histology and only 50% determined by MRI-PDFF assuming the absence of technical factors for the moment [53]. Second, hepatic fat content is a continuous variable and grading steatosis on an ordinal scale may introduce potential misclassification at the border zones between grades [24]. Tang et al [10] found that false-positive and false-negative results were prone to be at boundaries between histological grades. Third, the tissue volume sampled by biopsy is approximately 1/50,000 of the entire liver [54]. This inherently microscopic nature of biopsy can introduce sampling variability and misclassification especially in the liver with heterogeneous fat distribution. Our results together with a previous meta-analysis [33] suggest that MRI-PDFF has high accuracy on the assessment of LS with either histology or MRS-PDFF [33] as reference standard, and it has been shown that hepatic MRI-PDFF has high repeatability and reproducibility [33, 55]; additionally, MRI-PDFF is able to quantify the fat fraction in the whole liver noninvasively. Thus, MRI-PDFF may be used as an alternative to biopsy to quantify hepatic fat content continuously instead of grading on an ordinal scale in the future.

MRS-PDFF has been another accepted noninvasive reference standard in multiple studies focusing on the hepatic MRI-PDFF quantification. A patient-level meta-analysis assessing the linearity and bias between MRI- and MRS-PDFF for the quantification of LS has shown high accuracy of MRI-PDFF with MRS-PDFF as reference standard, which is complementary to our results that MRI-PDFF has high accuracy for the assessment of LS with histology as reference standard. Although MRS-PDFF utilizes a voxel that is about 8–27 cm^3 which is many times greater in volume than a typical liver biopsy specimen, the technique still samples a fraction of the liver. In addition, MRS only measures PDFF within the

pre-located voxel, and the location scan and acquisition scan are from different breath-holds, which is prone to errors in position. In contrast, regions of interest (ROIs) can be located directly on the MRI-PDFF map of the entire liver. MRS takes the same amount of acquisition time as MRI for measuring PDFF, and previous studies have shown hepatic MRS-PDFF may have slightly lower repeatability than MRI-PDFF [55, 56]. Therefore, together with our results, hepatic MRI-PDFF may be an appropriate alternative to MRS-PDFF.

Our study has several limitations. First, our meta-analysis was based on study-level rather than patient-level data, and the grading thresholds of MRI-PDFF varied across original studies. We attempted to obtain raw data from the authors of included studies, but were unable to do so. As a result, we were not able to derive thresholds of MRI-PDFF for the grading of LS based on the raw de-identified data. This drawback notwithstanding the practice of using study-level data is well-established in multiple published meta-analyses examining diagnostic performance [45, 57–59] and is consistent with the standards used in the Cochrane handbook for diagnostic systematic reviews [60]. To account for threshold differences among original studies, we conducted SROC curve analyses and regarded AUC as the indicator of accuracy. Second, the number of eligible studies was relatively few. In the future, more studies are needed to be included in updated meta-analyses for more comprehensive assessment. Third, although we performed a comprehensive literature search in several authoritative databases, we excluded non–English language articles and meeting abstracts, which might influence the results to some extent.

In conclusion, based on 13 eligible studies with various populations, MR parameters, and results, our meta-analysis provides comprehensive evidence that MRI-PDFF has high diagnostic accuracy at detecting and grading LS with histology as reference standard. These results suggest that MRI-PDFF is able to provide an accurate quantification of LS in clinical trials and patient care.

Funding This study has received funding by the National Natural Science Foundation of China, No. 81471658.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Bin Song.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was not required for this study because this study is based on the published studies to perform the meta-analysis.

Ethical approval Institutional Review Board approval was not required because this study is based on the published studies to perform data analysis.

Methodology
- retrospective
- diagnostic or prognostic study
- performed at one institution

References

1. Reeder SB, Cruite I, Hamilton G, Sirlin CB (2011) Quantitative assessment of liver fat with magnetic resonance imaging and spectroscopy. J Magn Reson Imaging 34:729–749
2. Rinella ME (2015) Nonalcoholic fatty liver disease: a systematic review. JAMA 313:2263–2273
3. Chalasani N, Younossi Z, Lavine JE et al (2012) The diagnosis and management of non-alcoholic fatty liver disease: practice guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. Hepatology 55:2005–2023
4. Stinton LM, Loomba R (2014) Recommendations for liver biopsy evaluation in non-alcoholic fatty liver disease. Minerva Gastroenterol Dietol 60:5–13
5. Fernandez-Salazar L, Velayos B, Aller R, Lozano F, Garrote JA, Gonzalez JM (2011) Percutaneous liver biopsy: patients' point of view. Scand J Gastroenterol 46:727–731
6. Bravo AA, Sheth SG, Chopra S (2001) Liver biopsy. N Engl J Med 344:495–500
7. Regev A, Berho M, Jeffers LJ et al (2002) Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. Am J Gastroenterol 97:2614–2618
8. Bedogni G, Nobili V, Tiribelli C (2014) Epidemiology of fatty liver: an update. World J Gastroenterol 20:9050–9054
9. Idilman IS, Keskin O, Celik A et al (2016) A comparison of liver fat content as determined by magnetic resonance imaging-proton density fat fraction and MRS versus liver histology in non-alcoholic fatty liver disease. Acta Radiol 57:271–278
10. Tang A, Desai A, Hamilton G et al (2015) Accuracy of MR imaging-estimated proton density fat fraction for classification of dichotomized histologic steatosis grades in nonalcoholic fatty liver disease. Radiology 274:416–425
11. Hamilton G, Yokoo T, Bydder M et al (2011) In vivo characterization of the liver fat ^{1}H MR spectrum. NMR Biomed 24:784–790
12. Meisamy S, Hines CDG, Hamilton G et al (2011) Quantification of hepatic steatosis with T1-independent, T2*-corrected MR imaging with spectral modeling of fat: blinded comparison with MR spectroscopy. Radiology 258:767–775
13. Kim KY, Song JS, Kannengiesser S, Han YM (2015) Hepatic fat quantification using the proton density fat fraction (PDFF): utility of free-drawn-PDFF with a large coverage area. Radiol Med 120:1083–1093
14. Motosugi U, Hernando D, Bannas P et al (2015) Quantification of liver fat with respiratory-gated quantitative chemical shift encoded MRI. J Magn Reson Imaging 42:1241–1248
15. Rehm JL, Wolfgram PM, Hernando D, Eickhoff JC, Allen DB, Reeder SB (2015) Proton density fat-fraction is an accurate biomarker of hepatic steatosis in adolescent girls and young women. Eur Radiol 25:2921–2930
16. Hetterich H, Bayerl C, Peters A et al (2016) Feasibility of a three-step magnetic resonance imaging approach for the assessment of hepatic steatosis in an asymptomatic study population. Eur Radiol 26:1895–1904

17. Park CC, Hooker C, Hooker JC et al (2018) Assessment of a high-SNR chemical-shift-encoded MRI with complex reconstruction for proton density fat fraction (PDFF) estimation overall and in the low-fat range. J Magn Reson Imaging. https://doi.org/10.1002/jmri.26168
18. Bonekamp S, Tang A, Mashhood A et al (2014) Spatial distribution of MRI-determined hepatic proton density fat fraction in adults with nonalcoholic fatty liver disease. J Magn Reson Imaging 39:1525–1532
19. Angulo P (2006) Nonalcoholic fatty liver disease and liver transplantation. Liver Transpl 12:523–534
20. Middleton MS, Van Natta ML, Heba ER et al (2018) Diagnostic accuracy of magnetic resonance imaging hepatic proton density fat fraction in pediatric nonalcoholic fatty liver disease. Hepatology 67:858–872
21. Park CC, Phirum N, Hernandez C et al (2017) 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 152:598–607
22. Paige JS, Bernstein GS, Heba E et al (2017) A pilo comparative study of quantitative ultrasound, conventional ultrasound, and MRI for predicting histology-determined steatosis grade in adult nonalcoholic fatty liver disease. AJR Am J Roentgenol 208:W168–W177
23. Middleton MS, Heba ER, Hooker CA et al (2017) Agreement between magnetic resonance imaging proton density fat fraction measurements and pathologist-assigned steatosis grades of liver biopsies from adults with nonalcoholic steatohepatitis. Gastroenterology 153:753–761
24. Schwimmer JB, Middleton MS, Behling C et al (2015) Magnetic resonance imaging and liver histology as biomarkers of hepatic steatosis in children with nonalcoholic fatty liver disease. Hepatology 61:1887–1895
25. Paparo F, Cenderello G, Revelli M et al (2015) Diagnostic value of MRI proton density fat fraction for assessing liver steatosis in chronic viral C hepatitis. Biomed Res Int 2015:758164
26. Chiang HJ, Lin LH, Li CW et al (2014) Magnetic resonance fat quantification in living donor liver transplantation. Transplant Proc 46:666–668
27. Tang A, Tan J, Sun M et al (2013) Nonalcoholic fatty liver disease: MR imaging of liver proton density fat fraction to assess hepatic steatosis. Radiology 267:422–431
28. Idilman IS, Aniktar H, Idilman R et al (2013) Hepatic steatosis: quantification by proton density fat fraction with MR imaging versus liver biopsy. Radiology 267:767–775
29. Kuhn JP, Hernando D, Munoz del Rio A et al (2012) Effect of multipeak spectral modeling of fat for liver iron and fat quantification: correlation of biopsy with MR imaging results. Radiology 265:133–142
30. Joe E, Lee JM, Kim KW et al (2012) Quantification of hepatic macrosteatosis in living, related liver donors using T1-independent, T2corrected chemical shift MRI. J Magn Reson Imaging 36:1124–1130
31. Zand KA, Shah A, Heba E et al (2015) Accuracy of multiecho magnitude-based MRI (M-MRI) for estimation of hepatic proton density fat fraction (PDFF) in children. J Magn Reson Imaging 42:1223–1232
32. Heba ER, Desai A, Zand KA et al (2016) Accuracy and the effect of possible subject-based confounders of magnitude-based MRI for estimating hepatic proton density fat fraction in adults, using MR spectroscopy as reference. J Magn Reson Imaging 43:398–406
33. Yokoo T, Serai SD, Pirasteh A et al (2018) Linearity, bias, and precision of hepatic proton density fat fraction measurements by using MR imaging: a meta-analysis. Radiology 286:486–498
34. Whiting PF, Rutjes AW, Westwood ME et al (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 155:529–536
35. Kleiner DE, Brunt EM, Van Natta M et al (2005) Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 41:1313–1321
36. Higgins JPT, Thompson SG, Deeks JJ, DG A (2003) Mearuing inconsistency in meta-analyses. BMJ 327:557–560
37. Dwamena BA (2007) Midas: a program for meta-analytical integration of diagnostic accuracy studies in Stata. Division of Nuclear Medicine, Department of Radiology, University of Michigan Medical School, Ann Arbor, Michigan. Available via http://fmwww.bc.edu/repec/bocode/m/midas.html. Accessed 16 June 2018
38. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J (2003) How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess Rep 7:1–76
39. Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A (2006) Meta-DiSc: a software for meta-analysis of test accuracy data. BMC Med Res Methodol 6:31
40. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 58:982–990
41. Cronin P, Kelly AM, Altaee D, Foerster B, Petrou M, Dwamena BA (2018) How to perform a systematic review and meta-analysis of diagnostic imaging studies. Acad Radiol 25:573–593
42. Moses LE, Shapiro D, Littenberg B (1993) Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med 12:1293–1316
43. Dukic V, Gatsonis C (2003) Meta-analysis of diagnostic test accuracy assessment studies with varying number of thresholds. Biometrics 59:936–946
44. Deeks JJ, Macaskill P, Irwig L (2005) The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 58:882–893
45. Bohte AE, van Werven JR, Bipat S, Stoker J (2011) The diagnostic accuracy of US, CT, MRI and 1H-MRS for the evaluation of hepatic steatosis compared with liver biopsy: a meta-analysis. Eur Radiol 21:87–97
46. Deng J, Fishbein MH, Rigsby CK, Zhang G, Schoeneman SE, Donaldson JS (2014) Quantitative MRI for hepatic fat fraction and T2* measurement in pediatric patients with non-alcoholic fatty liver disease. Pediatr Radiol 44:1379–1387
47. Idilman IS, Keskin O, Elhan AH, Idilman R, Karcaaltincaba M (2014) Impact of sequential proton density fat fraction for quantification of hepatic steatosis in nonalcoholic fatty liver disease. Scand J Gastroenterol 49:617–624
48. Joy TR, McKenzie CA, Tirona RG et al (2017) Sitagliptin in patients with non-alcoholic steatohepatitis: a randomized, placebo-controlled trial. World J Gastroenterol 23:141–150
49. Kukuk GM, Hittatiya K, Sprinkart AM et al (2015) Comparison between modified Dixon MRI techniques, MR spectroscopic relaxometry, and different histologic quantification methods in the assessment of hepatic steatosis. Eur Radiol 25:2869–2879
50. Loomba R, Sirlin CB, Ang B et al (2015) Ezetimibe for the treatment of nonalcoholic steatohepatitis: assessment by novel magnetic resonance imaging and magnetic resonance elastography in a randomized trial (MOZART trial). Hepatology 61:1239–1250
51. Cui J, Philo L, Nguyen P et al (2016) Sitagliptin vs. placebo for non-alcoholic fatty liver disease: a randomized controlled trial. J Hepatol 65:369–376

52. Caussy C, Alquiraish MH, Nguyen P et al (2018) Optimal threshold of controlled attenuation parameter with MRI-PDFF as the gold standard for the detection of hepatic steatosis. Hepatology 67: 1348–1359
53. Jayakumar S, Middleton MS, Lawitz EJ et al (2019) Longitudinal correlations between MRE, MRI-PDFF, and liver histology in patients with non-alcoholic steatohepatitis: analysis of data from a phase II trial of selonsertib. J Hepatol 70:133–141
54. Ratziu V, Charlotte F, Heurtier A et al (2005) Sampling variability of liver biopsy in nonalcoholic fatty liver disease. Gastroenterology 128:1898–1906
55. Tyagi A, Yeganeh O, Levin Y et al (2015) Intra- and inter-examination repeatability of magnetic resonance spectroscopy, magnitude-based MRI, and complex-based MRI for estimation of hepatic proton density fat fraction in overweight and obese children and adults. Abdom Imaging 40:3070–3077
56. Yokoo T, Shiehmorteza M, Hamilton G et al (2011) Estimation of hepatic proton-density fat fraction by using MR imaging at 3.0 T. Radiology 258:749–759
57. Tian F, Shen G, Deng Y, Diao W, Jia Z (2017) The accuracy of 18F-FDG PET/CT in predicting the pathological response to neoadjuvant chemotherapy in patients with breast cancer: a meta-analysis and systematic review. Eur Radiol 4786–4796. https://doi.org/10.1007/s00330-017-4831-y
58. Hernaez R, Lazo M, Bonekamp S et al (2011) Diagnostic accuracy and reliability of ultrasonography for the detection of fatty liver: a meta-analysis. Hepatology 54:1082–1090
59. Foerster BR, Dwamena BA, Petrou M, Carlos RC, Callaghan BC, Pomper MG (2012) Diagnostic accuracy using diffusion tensor imaging in the diagnosis of ALS: a meta-analysis. Acad Radiol 19:1075–1086
60. Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y (2010) Chapter 10: Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C (eds) Cochrane handbook for systematic reviews of diagnostic test accuracy Version 1.0. The Cochrane Collaboration. Available via http://srdta.cochrane.org/. Accessed 21 June 2018