Int J Clin Exp Med 2019;12(1):1186-1195
www.ijcem.com /ISSN:1940-5901/IJCEM0072788

Original Article

Diagnostic accuracy of magnetic resonance imaging-proton density fat fraction in quantifying liver steatosis: a systematic review and meta-analysis

Yan Chen[1*], Li Zhang[2*], Fuqian He[1], Gongxiang Liu[1], Kun Yang[1], Yan Tie[3], Huatian Gan[1,4]

[1]The Center of Gerontology and Geriatrics, West China Hospital, Sichuan University, Chengdu, Sichuan, China; [2]Department of Elderly Digestive, Sichuan Provincial People's Hospital, Chengdu, Sichuan, China; Departments of [3]Oncology, [4]Gastroenterology, West China Hospital, Sichuan University, Chengdu, Sichuan, China. [*]Equal contributors.

Received January 15, 2018; Accepted October 7, 2018; Epub January 15, 2019; Published January 30, 2019

Abstract: Objectives: Magnetic resonance imaging-proton density fat fraction (MRI-PDFF) is an imaging technique which is used to estimate liver steatosis. We performed a meta-analysis to assess the diagnostic accuracy of MRI-PDFF in quantifying liver steatosis, with histopathology as the reference standard. Methods: We searched PubMed, Medline, Embase and Cochrane databases and divided the steatosis grades of biopsies into four grade (grade 0, 0-5%; grade 1, 5%-33%; grade 2, 33%-66%; grade 3, >66%). Area under receiver operating characteristic curves (AUC of SROC), sensitivity, specificity, positive and negative likelihood ratio (LR+ and LR-), and diagnostic odds ratio (DOR) were calculated. Results: We included 11 studies with 653 eligible patients undergoing both liver biopsy and MRI-PDFF. There are only sufficient data for grade 0 and grade 2 estimation. For grade 0 sensitivity, specificity and AUC of SROC were 0.91 (0.87, 0.94), 0.88 (0.82, 0.92), and 0.9594 (SE = 0.0139), respectively. For grade 2 sensitivity, specificity and AUC of SROC were 0.79 (0.72, 0.85), 0.88 (0.82, 0.92) and 0.9298 (SE = 0.0195), respectively. Conclusions: Our meta-analysis shows that MRI-PDFF technique has high accuracy in diagnosing and quantifying liver steatosis, and may provide a clear, accessible, and accurate quantification of liver fat content.

Keywords: MRI-PDFF, liver steatosis, diagnostic accuracy, meta-analysis

Introduction

Liver steatosis is increasing rapidly worldwide, largely as a result of its association with obesity and insulin resistance in non-alcoholic fatty liver disease (NAFLD) [1-3]. Liver steatosis is defined pathologically as abnormal and excessive intracellular accumulation of fat, primarily as triglycerides, in hepatocytes [4]. The clinical manifestation of liver steatosis ranges from non-alcoholic fatty liver disease (NAFLD) to non-alcoholic steatohepatitis (NASH). The latter condition can progress into fatty liver cirrhosis and hepatocellular carcinoma [4, 5]. Hepatic steatosis is also associated with alcoholic liver disease, viral hepatitis, human immunodeficiency virus (HIV), genetic lipodystrophies, cystic fibrosis liver disease, and hepatotoxicity from various therapeutic agents.

Liver biopsy is the currently recommended "gold standard" method for the diagnosis and grading of steatosis [6, 7]. However, it is an invasive procedure that is associated with a significant risk of complications and sampling error due to small sample size [6, 8]. Although magnetic resonance spectroscopy (MRS) generally is recommended as the most accurate imaging technique [9] among the non-invasive methods to evaluate liver fat content, and is available in nearly all clinical MRI systems, it also has some drawbacks. First, it cannot cover the entire liver (covering a volume of approximately 4 cm^3). Second, diagnostic radiology users may not be proficient in or comfortable with its use compared with the PDFF imaging method, which more closely resembles commonly used techniques [10].

The proton density fat fraction (PDFF, the fraction of the liver proton density attributable to liver fat) measurement is a recently developed advanced MRI-based technique for quantitative assessment of hepatic steatosis. MRI-PDFF

addresses confounding factors such as T1 bias, T2* decay, spectral complexity of fat, noise bias, and eddy currents in conventional MRI [11, 12]. In contrast to spectroscopy, novel reconstruction approaches can be used to substantially reduce the length of the patient breath holding required to collect PDFF measurements [13, 14]. M*ost importantly, MRI-PDFF shows good correlation with the grade of* steatosis in the liver both by biopsy [15-18] and by MRS [19-23].

Several previous studies have assessed the diagnostic accuracy of MRI-PDFF for assessing liver steatosis, but some of them have small sample sizes and, to some extent, different quantitative diagnosis and grading results. Therefore, the present meta-analysis aims to quantify the diagnostic accuracy of MRI-PDFF for liver steatosis, using histopathology as reference standard.

Methods

Literature search

A thorough search of PubMed, Medline, Embase and Cochrane databases, was conducted for relevant publications evaluating the diagnostic accuracy of MRI-PDFF for the diagnosis and grading of liver steatosis in humans, using histopathology as reference standard, until Nov 16, 2017. We used Medical Subject Headings (MeSH) terms and free text keywords: NAFLD, non-alcoholic fatty liver disease, liver or hepatic fat fraction, liver or hepatic fat content, steatosis, steatohepatitis, hepatosteatosis, NASH, non-alcoholic steatohepatitis. These search terms were combined with: magnetic resonance imaging, MRI, proton density fat fraction, and PDFF. There were no language and publication date restrictions. The reference lists of obtained studies and reviews were screened to select potentially relevant articles. Ethical approval was not required.

Study inclusion

Initially, two reviewers read the titles and abstracts of all the obtained studies to identify potentially relevant articles. Then, they independently reviewed the full texts of potentially relevant articles for inclusion.

Inclusion criteria were as following: (1) studies assessing MRI-PDFF for the diagnosis and quantification of liver steatosis; (2) studies using histopathology as the reference test; (3) studies enrolling ≥10 human individuals; (4) studies with data of diagnostic accuracy reported, including at least sensitivity, specificity, and AUROC or these parameters could be calculated.

Exclusion criteria were as following: (1) non-human studies; (2) studies that were not original research, letters, case reports, editorials or reviews; (3) studies reporting combined data from different imaging methods, or studies of which data on the diagnostic accuracy of MRI-PDFF could not be calculated; (4) duplicate publications.

Data extraction

Two reviewers independently extracted the following data: (1) authors and publication time, country, study design (prospective/retrospective study), sample size, male-female ratio, age, body mass index (BMI), patient spectrum; (2) MR imaging and liver biopsy features; (3) MRI-PDFF diagnostic accuracy relevant data, including sensitivity, specificity, and AUROC, cut-off, true positive (TP), false positive (FP), false negative (FN), true negative (TN). When disagreement arose, resolution was reached through face-to-face discussion between the two reviewers.

Quality assessment

The quality of diagnostic accuracy studies (QUADAS-2) [24] was applied to evaluate the quality of the included studies. If the answers to all signalling questions for a domain were "yes", then the risk of bias was judged as being low. If any answer to the signalling questions was "no", then the potential for bias existed. Then, the reviewer judged the risk of bias. The "unclear" category was used only when insufficient data were reported to permit a judgement. Specifically, for each category (risk of bias and applicability concerns), studies with ≥2 domains of high risk were designated high risk; those with only one domain of high risk would be designated medium risk; those with no domain of high risk would be designated low risk.

Statistical analysis

Data were extracted and entered into MetaDisc 1.4 to analyse forest plots and calculate AUC of

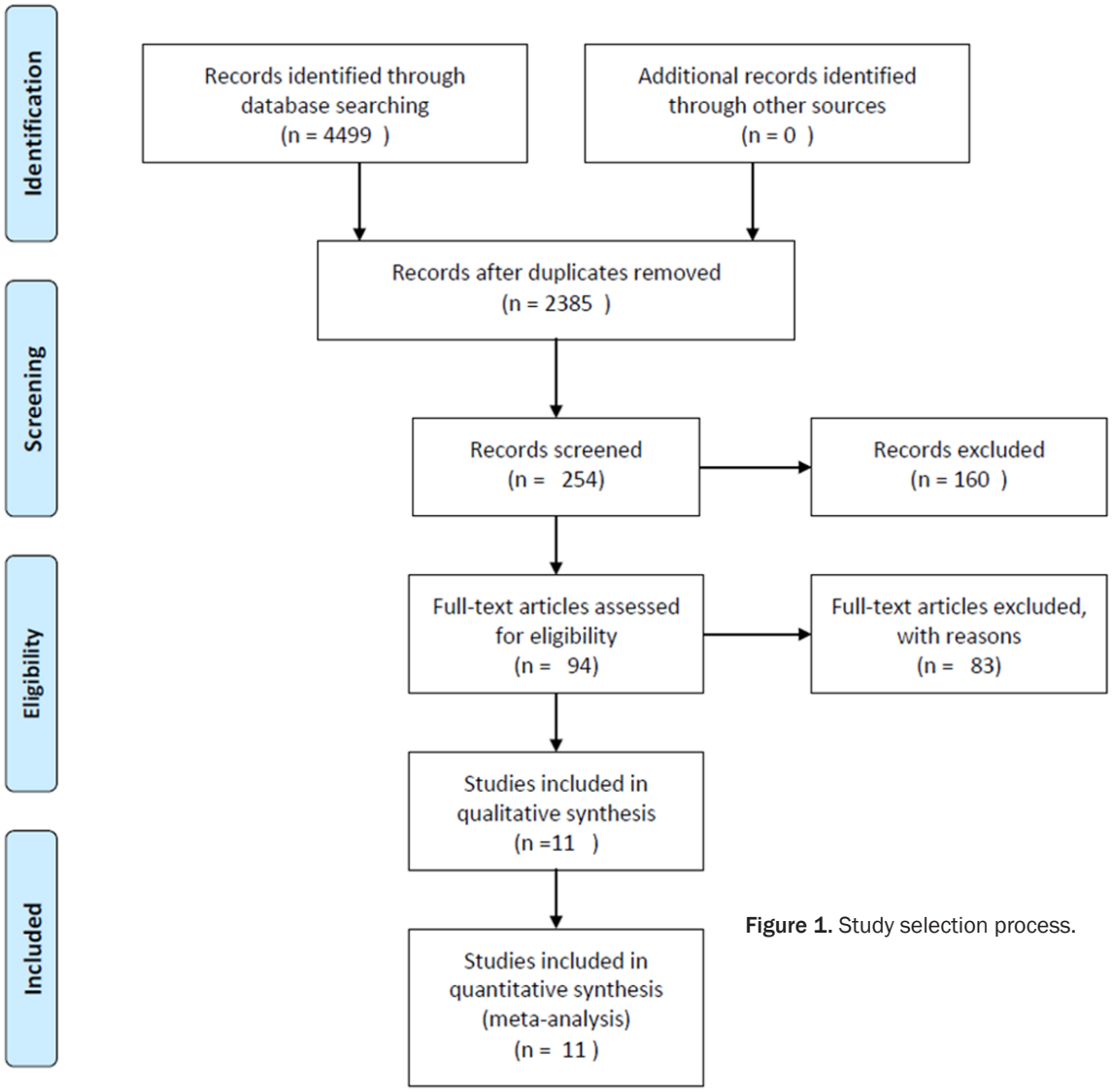

Figure 1. Study selection process.

SROC, summary statistics for sensitivity, specificity, positive and negative likelihood ratio (LR+/LR-), and diagnostic odds ratio (DOR) with 95% confidence intervals (CIs). Heterogeneity across studies was tested using an I^2 test. An I^2 statistic greater than 50% was considered statistically significant heterogeneity. If substantial heterogeneity existed, the random effects model (the DerSimonian - Laird method) was used. If not, the fixed effects model (the Mantel-Haenszel method) was applied. Egger's test was performed by STATA to investigate the publication bias, and a *p*-value less than 0.1 was set to reflect statistically significant publication bias. Because different study used slightly different threshold values for steatosis grading on liver biopsy, so we grouped these threshold values for liver biopsies into four groups, basically defined with the NASH nonalcoholic steatohepatitis CRN Clinical Research Network scoring system, to perform the meta-analysis: Grade 0: steatosis on biopsy threshold values: 0-5%; Grade 1: steatosis on biopsy threshold values: 5%-33%; Grade 2: steatosis on biopsy threshold values: 33%-66%; Grade 3: steatosis on biopsy threshold values: >66%.

Results

Description of included studies

A total of 4499 articles were identified at the initial search. After careful examination of the titles and abstracts, 4245 articles were exclud-

Table 1. Characteristics of included studies

Study	Country	Design	Patient, n (m/f)	Age, yrs, mean ± SD (range)	BMI, kg/m^2 (range)	Patient spectrum, n	Technique
Rinella 2003	US	P	22 (13/9)	Liver donors: 37 (23, 52) NAFLD: 37 (22, 50)	Liver donors: 31 (26, 35.5) NAFLD: 35 (28.2, 40.8)	15 living liver donors; 7 NAFLD	MRI
Yoshimitsu 2008	Japan	R	58 (35/23)	36.5 (20, 61)	25.1 ± 0.7 (18.3, 32.6)	38 potential living liver donors; 20 liver metastases	MRI,CT
d'Assignies 2009	Canada	P	20 (15/5)	53.7 ± 11.1 (33,72)	27.7 ± 3.9	14 NAFLD, 6 AFLD	MRI, MRS, DWI
Kang 2012	Korea	P	56 (35/21)	55 ± 10.1 (31-75)	24.5 ± 3.24 (15.2-33.2)	56 hepatic masse and chronic liver diseases	MRI, MRS
Benjamin 2013	Austria	R	31 (25/6)	50 (30, 72)	NR	31suspected hepatic iron overload or diffuse liver disease	MRI
Tang 2013	US	P	77 (61/16)	14 (8-61)	Adults: 33.2 ± 6.0 Children: 2.3 ± 0.4	77 NAFLD	MRI
Idilman 2013	Turkey	R	70 (40/30)	44.7 ± 13.1 (16–69)	29.9 ± 4.3	70 NAFLD	MRI
Tang 2015	US	P	89 (38/51)	51.0 ± 13.0 (22–80)	30.6 ± 5.0	89 NAFLD	MRI
Paparo 2015	Italy	P	77 (43/34)	51.31 ± 11.27 (18-81)	22.39 ± 2.27	77 HS with untreated chronic viral hepatitis C.	MRI
Idilman 2016	Turkey	R	19 (15/4)	41.7 ± 13.1	27.5 ± 3.3	19 NAFLD	MRI, MRS
Imajo 2016	Japan	P	152 (87/65)	NAFLD: 57.5 ± 14.6 Control: 52.1 ± 15.1	NAFLD: 21.9 ± 0.69 Control: 28.1 ± 4.63	142 NAFLD; 10 without NAFLD	MRI, TE-CAP

P, prospective study; R, retrospective study; M, male; F, female; SD, standard deviation; BMI,body mass index; n, number; NAFLD, Non-alcoholic Fatty Liver Disease; HS, hepatic steatosis; MRS, MR spectroscopy; DMI, diffusion-weighted MR imaging; TE-CAP, Transient elastography-based controlled attenuation parameter.

Table 2. Quality assessment

Study	Risk of bias					Applicability concerns			
	Patient selection	Index test	Reference	Flow & timing	Total	Patient selection	Index test	Reference	Total
Rinella 2003	H	L	L	L	M	H	L	U	M
Yoshimitsu 2008	H	U	U	M	L	H	L	L	M
d'Assignies 2009	H	L	L	L	M	L	L	L	L
Kang 2012	L	L	L	L	L	L	L	L	L
Benjamin 2013	H	U	U	L	M	H	L	L	M
Tang 2013	L	L	L	L	L	L	L	L	L
Idilman 2013	L	L	L	L	L	L	L	H	M
Tang 2015	L	L	L	L	L	L	L	L	L
Paparo 2015	L	L	L	L	L	H	L	L	M
Idilman 2016	L	L	L	L	L	L	L	H	M
Imajo 2016	L	L	L	L	L	L	L	L	L

H, high risk; M, Medium risk, L, low risk; U, unclear risk.

Table 3. Statistical summary of outcomes

Steatosis grade	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	DOR (95% CI)	AUC of SROC (SE)
Grade 0 (0-5%)	0.91 (0.87, 0.94)	0.88 (0.82, 0.92)	5.79 (3.94, 8.50)	0.12 (0.08, 0.18)	74.93 (32.64, 171.99)	0.9594 (0.0139)
Grade 2 (5%-33%)	0.79 (0.72, 0.85)	0.88 (0.82, 0.92)	5.07 (2.92, 8.80)	0.22 (0.11, 0.43)	35.56 (16.84, 75.09)	0.9298 (0.0195)

LR+, positive likelihood ratio; LR-, negative likelihood ratio; DOR, diagnostic odds ratio; AUC of SROC, area under receiver operating characteristic curves; SE, standard error.

Table 4. Number of studies and patients in each group analyzed

Group	Steatosis grade	n. (sensitivity)	n. (specificity)	n. (LR+)	n. (LR-)	n. (DOR)	n. (AUC of SROC)
Grade 0	0-5%	8 (453)	8 (453)	8 (453)	8 (453)	8 (453)	8 (453)
Grade 1	5%-33%	1	-	-	-	-	-
Grade 2	33%-66%	6 (453)	6 (453)	6 (453)	6 (453)	6 (453)	6 (453)
Grade 3	>66%	2	-	-	-	-	-

n, number of studies and patients (in parentheses).

ed for having no or minimal relevance to the accurate diagnosis of liver steatosis by MRI-PDFF. The full-text of 254 articles was retrieved. After careful reading of these full texts, 94 articles were retained. Of these, 18 articles used MRS as a reference, 39 reported other methods of diagnosis (such as correlation between histological assessment of steatosis by MRS and MRI and reproducibility of MRI-PDFF), but were not relevant to diagnosis accuracy, 18 were reviews or editorials, and 8 had incomplete result data relevant to diagnostic accuracy. These articles were excluded, leaving 11 eligible articles [15-18, 25-31], which were included in this meta-analysis. The study selection process is summarized in **Figure 1**. The details of the search strategy are reported in Table S1. Publication bias, as assessed by Egger's test, was not found, with P = 0.145.

A total of 653 eligible patients undergoing both liver biopsy and MRI-PDFF were included. Seven studies were prospective [15, 17, 18, 25, 27, 30, 31], and 4 were retrospective [16, 26, 28, 29]. Among these, publication years ranged from 2003 to 2016, sample sizes ranged from 19 to 152, and the most frequent associated disease was NAFLD/HS (495/653). The detailed characteristics of the included studies are listed in **Table 1**. Number of studies and patients in each group analyzed are presented in **Table 4**.

MR imaging and liver biopsy features are summarized in Table S2. All included studies used both 1.5 Tesla and 3.0 Tesla magnetic field strengths. The sequences used included T1-weighted dual gradient echo, T1-weighted spoiled gradient dual echo, two-dimensional transverse spoiled gradient-echo, T1 fast low-

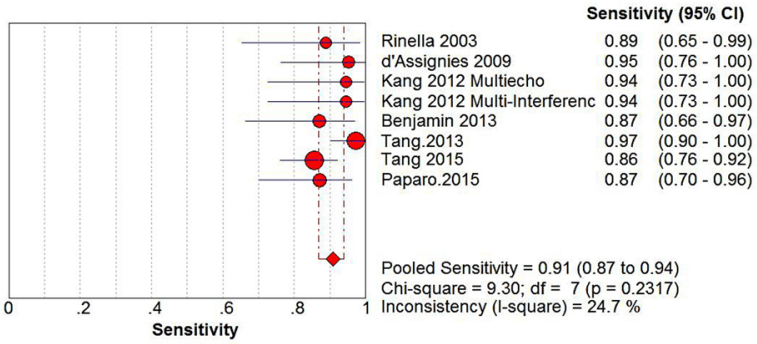

Figure 2. Sensitivity for Grade 0.

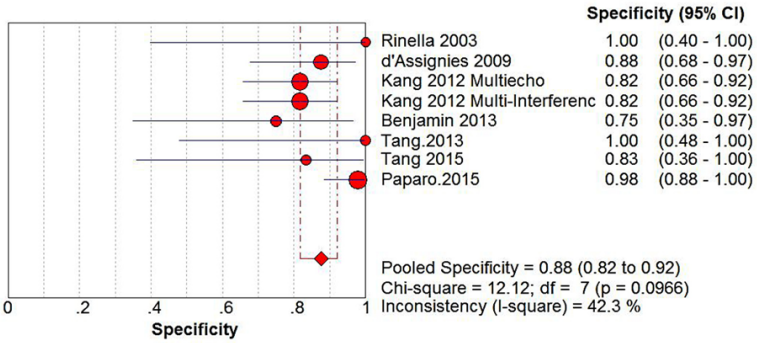

Figure 3. Specificity for Grade 0.

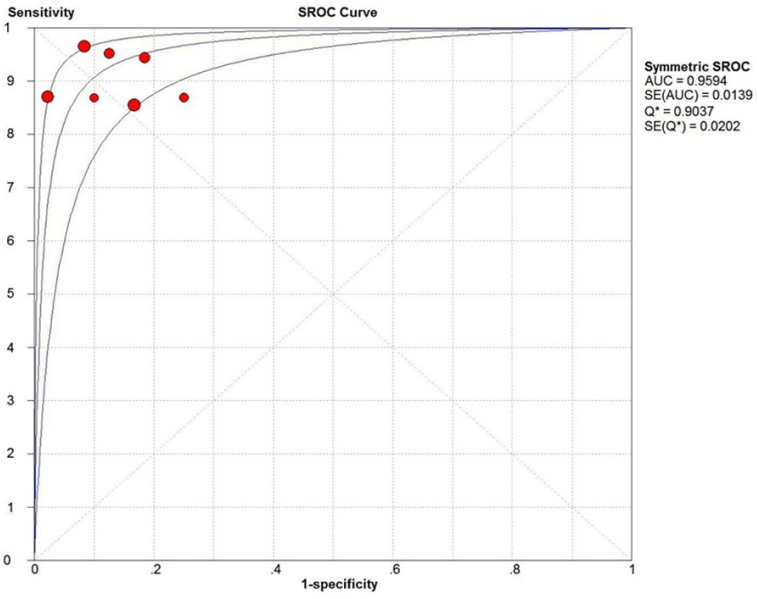

Figure 4. AUC of SROC for Grade 0.

angle shot, T1-independent multi-echo gradient-echo, two-dimensional spoiled gradient echo, multi-echo and multi-interference. Nine of the eleven included studies have reported a correction for T2* effects and breath holding. The data for calculation of MRI-PDFF diagnosis accuracy, including true positive (TP), false positive (FP), false negative (FN), true negative (TN), sensitivity, specificity, AUROC, and cut-off, were extracted and summarized in Table S3.

Quality assessment

According to QUADAS-2 (A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies), the quality assessment composed of two categories: risk of bias and applicability concerns. In general, risk of bias and applicability of the included studies were low to medium. For risk of bias, we included 8 low-risk studies and 3 medium-risk studies. For applicability concerns, we included 5 low-risk studies and 6 medium-risk studies (**Table 2**).

Statistical pooling of outcomes

A summary statistics for sensitivity, specificity, positive and negative likelihood ratio (LR+/LR-), DOR, AUC of SROC with 95% confidence intervals (CIs) are showed in **Table 3**.

Grade 0 (0-5%): Given that no significant heterogeneity existed for sensitivity, specificity, LR+, LR-, and DOR (I^2 = 24.7%, 42.3%, 0.0%, 8.5%, 0.0%, respectively), Mantel-Haenszel fixed effect model were used. As displayed in **Figures 2** and **3**, our meta-analysis showed that the sensitivity and specificity were 0.91 (0.87, 0.94) and 0.88 (0.82, 0.92), respectively. LR+ and LR- was 5.79 (3.94,

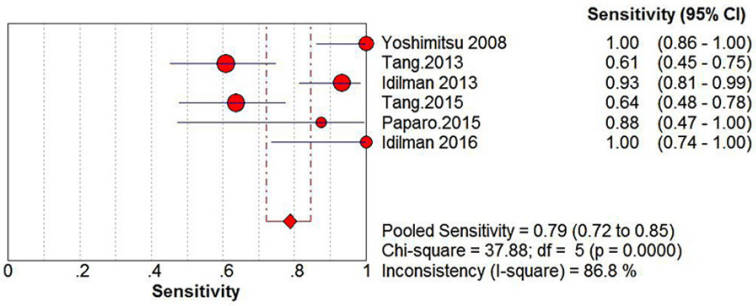

Figure 5. Sensitivity for Grade 2.

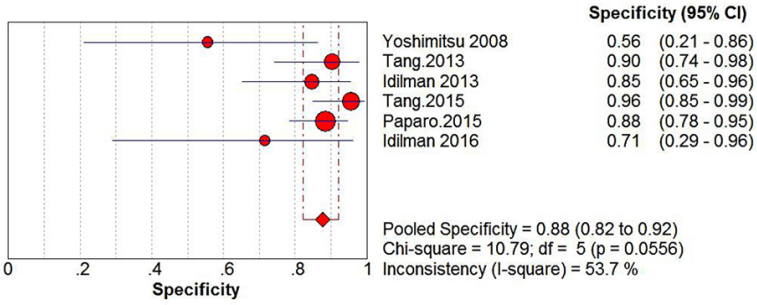

Figure 6. Specificity for Grade 2.

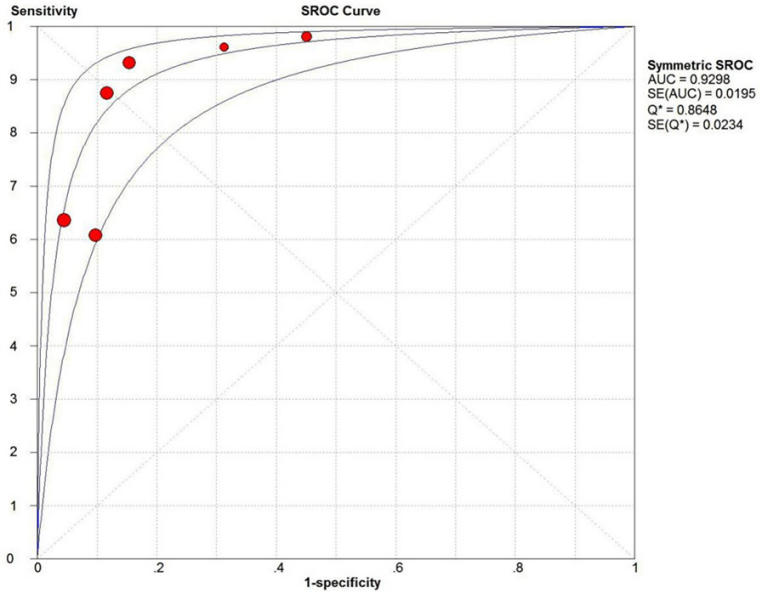

Figure 7. AUC of SROC for Grade 2.

8.50) and 0.12 (0.08, 0.18), respectively. DOR was 74.93 (32.64, 171.99). AUC of SROC was 0.9594 (standard error [SE] = 0.0139) (**Figure 4**).

Grade 2 (33%-66%): There was heterogeneity in this group for sensitivity, specificity, LR+ and LR- (I^2 =86.8%, 53.7%, 53.5%, 71.2%, respectively). Except for DOR (I^2 = 0.0%). So DerSimonian-Laird random effects model was used. As displayed in **Figures 5** and **6**, our meta-analysis showed that sensitivity and specificity were 0.79 (0.72, 0.85) and 0.88 (0.82, 0.92), respectively. LR+ and LR- were 5.07 (2.92, 8.80) and 0.22 (0.11, 0.43), respectively. DOR was 35.56 (16.84, 75.09). AUC of SROC was 0.9298 (standard error [SE] = 0.0195) (**Figure 7**).

Discussion

Currently, liver biopsy remains reference standard to evaluate the grade of hepatic steatosis. However, biopsy has some limitations, and is not optimal for screening, monitoring, or clinical decision-making, and is not well suited to certain research studies. Therefore, a noninvasive and objective quantification of liver fat content may be preferable to liver biopsy in both clinical practice and research. Several noninvasive imaging techniques, including ultrasound, CT, MRS, and MRI, have been used to evaluate liver fat content. Ultrasound is the most common imaging method used to evaluate hepatic steatosis because of its low cost, its safety, and its availability. However, it has limited sensitivity, specificity, reproducibility [32, 33], and is challenging to perform in obese patients [33]. CT is another imaging modality, but its analysis can be confounded by several factors, and it cannot quantify steatosis beyond grading [34]. Furthermore, since CT relies on ionizing radiation [35], it is not suitable for use in children and in follow-up evaluations. Chemical shift based water-fat separation methods have evolved dramatically in their ability to quantify hepatic

steatosis in recent years. MRI-PDFF is an important MRI based accurate imaging modality. MRI-PDFF eliminates confounding factors and quantifies the fat fraction of the entire liver, which is limited in both MRS and liver biopsy. It would be of great value to evaluate the diagnostic accuracy of MRI-PDFF in grading liver steatosis for both clinical physicians and reseachers.

In this meta-analysis, we included 11 independent studies, with 653 patients undergoing both liver biopsy and MRI-PDFF. We divided liver steatosis into four grades according to NASH nonalcoholic steatohepatitis CRN Clinical Research Network scoring system.

Our result showed that, for grade 0, sensitivity, specificity and DOR were 0.91 (0.87, 0.94), 0.88 (0.82, 0.92), and 74.93 (32.64, 171.99), respectively. For grade 2, sensitivity, specificity and DOR were 0.79 (0.72, 0.85), 0.88 (0.82, 0.92), and 35.56 (16.84, 75.09), respectively. Our results showed that MRI-PDFF can provide high accuracy for diagnosis and quantifying liver steatosis.

SROC is proposed as an index to assess diagnostic accuracy in meta-analyses [36, 37]. In practice, AUC of SROC has a range from 0.5 (no better than chance) to 1 (a perfect test) [38]. In our review, AUC of SROC in grade 0 was 0.9594 (SE= 0.0139) and AUC of SROC in grade 2 was 0.9298 (SE = 0.0195), which suggest a high accuracy of MRI-PDFF in diagnosing and quantifying liver steatosis.

LR+ greater than 5 or LR- less than 0.2 can provide strong diagnostic evidence [39]. In our review, for grade 0, LR+ and LR- were 5.79 (3.94, 8.50) and 0.12 (0.08, 0.18), respectively. For grade 2, LR+ and LR- were 5.07 (2.92, 8.80) and 0.22 (0.11, 0.43), respectively, which also indicate a strong diagnostic evidence.

There are several limitations in our study. First, even though we excluded studies enrolling fewer than 10 individuals, some included studies still had small sample sizes because we only included studies using live biopsy as the reference stander. Second, we included a diverse spectrum of patients, and this diversity may have induced greater heterogeneity in the meta-analysis for grade 2. Third, because there were different grading standers of hepatic steatosis in different studies, so we divided the nearby cut-off values into one grade based on to NASH nonalcoholic steatohepatitis CRN Clinical Research Network scoring system. The ideal situation would be to analyze accuracy results for each cut-off value separately. Fourth, we only performed meta-analysis in only grade 0 and grade 2, because grade 1 and grade 3 had no sufficient data. In summary, our meta-analysis suggests that MRI-PDFF technique has high accuracy in diagnosing and quantifying liver steatosis. This method provides an accessible, accurate quantification of liver fat content without confounding factors and is generally preferable to MRS and liver biopsy for spatial coverage of the entire liver.

Acknowledgements

We acknowledge the work of Gongxiang Liu involved in data analysis.

Disclosure of conflict of interest

None.

Address correspondence to: Huatian Gan, The Center of Gerontology and Geriatrics, Department of Gastroenterology, West China Hospital, Sichuan University, No. 37 Guoxuexiang, Wuhou District, Chengdu, China. Tel: +86 18980601334; E-mail: ganhuatian123@163.com

References

[1] Angulo P. Nonalcoholic fatty liver disease. N Engl J Med 2002; 346: 1221-1231.
[2] Chavez-Tapia NC, Mendez-Sanchez N and Uribe M. The metabolic syndrome as a predictor of nonalcoholic fatty liver disease. Ann Intern Med 2006; 144: 379.
[3] Williams R. Global challenges in liver disease. Hepatology 2006; 44: 521-526.
[4] Day CP and Saksena S. Non-alcoholic steatohepatitis: definitions and pathogenesis. J Gastroenterol Hepatol 2002; 17 Suppl 3: S377-384.
[5] Harrison SA, Torgerson S and Hayashi PH. The natural history of nonalcoholic fatty liver disease: a clinical histopathological study. Am J Gastroenterol 2003; 98: 2042-2047.
[6] Cadranel JF. Good clinical practice guidelines for fine needle aspiration biopsy of the liver: past, present and future. Gastroenterol Clin Biol 2002; 26: 823-824.
[7] Chalasani N, Younossi Z, Lavine JE, Diehl AM, Brunt EM, Cusi K, Charlton M and Sanyal AJ.

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 2012; 55: 2005-2023.

[8] Bravo AA, Sheth SG and Chopra S. Liver biopsy. N Engl J Med 2001; 344: 495-500.

[9] Szczepaniak LS. Magnetic resonance spectroscopy to measure hepatic triglyceride content: prevalence of hepatic steatosis in the general population. Am J Physiol Endocrinol Metab 2005; 288: E462-E468.

[10] Meisamy S, Hines CD, Hamilton G, Sirlin CB, McKenzie CA, Yu H, Brittain JH and Reeder SB. Quantification of hepatic steatosis with T1-independent, T2-corrected MR imaging with spectral modeling of fat: blinded comparison with MR spectroscopy. Radiology 2011; 258: 767-775.

[11] Guiu B, Loffroy R, Petit JM, Aho S, Ben Salem D, Masson D, Hillon P, Cercueil JP, Krause D. Mapping of liver fat with triple-echo gradient echo imaging: validation against 3.0-T proton MR spectroscopy. Eur Radiol 2009; 19: 1786-93.

[12] Yokoo T, Bydder M, Hamilton G, Middleton MS, Gamst AC, Wolfson T, Hassanein T, Patton HM, Lavine JE, Schwimmer JB, Sirlin CB. Nonalcoholic fatty liver disease: diagnostic and fat-grading accuracy of low-flip-angle multiecho gradient-recalled-echo MR imaging at 1.5 T. Radiology 2009; 251: 67-76.

[13] Mann LW, Higgins DM, Peters CN, Cassidy S, Hodson KK, Coombs A, Taylor R and hollingsworth KG. accelerating MR imaging liver steatosis measurement using combined compressed sensing and parallel imaging: a quantitative evaluation. Radiology 2016; 278: 247-56.

[14] Sharma SD, Hu HH, Nayak KS. Accelerated T2*-compensated fat fraction quantification using a joint parallel imaging and compressed sensing framework. J Magn Reson Imaging 2013; 38: 1267-75.

[15] Tang A, Tan J, Sun M, Hamilton G, Bydder M, Wolfson T, Gamst AC, Middleton M, Brunt EM, Loomba R, Lavine JE, Schwimmer JB, Sirlin CB. Nonalcoholic fatty liver disease: MR imaging of liver proton density fat fraction to assess hepatic steatosis. Radiology 2013; 267: 422-31.

[16] Idilman IS, Aniktar H, Idilman R, Kabacam G, Savas B, Elhan A, Celik A, Bahar K, Karcaaltincaba M. Hepatic steatosis: quantification by proton density fat fraction with MR imaging versus liver biopsy. Radiology 2013; 267: 767-75.

[17] Tang A, Desai A, Hamilton G, Wolfson T, Gamst A, Lam J, Clark L, Hooker J, Chavez T, Ang BD, Middleton MS, Peterson M, Loomba R, Sirlin CB. Accuracy of MR imaging-estimated proton density fat fraction for classification of dichotomized histologic steatosis grades in nonalcoholic fatty liver disease. Radiology 2015; 274: 416-25.

[18] Paparo F, Cenderello G, Revelli M, Bacigalupo L, Rutigliani M, Zefiro D, Cevasco L, Amico M, Bandelloni R, Cassola G, Forni GL and Rollandi GA. Diagnostic value of MRI proton density fat fraction for assessing liver steatosis in chronic viral C hepatitis. Biomed Res Int 2015; 2015: 758164.

[19] Yokoo T, Bydder M, Hamilton G, Middleton MS, Gamst AC, Wolfson T, Hassanein T, Patton HM, Lavine JE, Schwimmer JB, Sirlin CB. Nonalcoholic fatty liver disease: diagnostic and fat-grading accuracy of low-flip-angle multiecho gradient-recalled-echo MR imaging at 1.5 T. Radiology 2009; 251: 67-76.

[20] Guiu B, Loffroy R, Petit JM, Aho S, Ben Salem D, Masson D, Hillon P, Cercueil JP and Krause D. Mapping of liver fat with triple-echo gradient echo imaging: validation against 3.0-T proton MR spectroscopy. Eur Radiol 2009; 19: 1786-1793.

[21] Hines CD, Frydrychowicz A, Hamilton G, Tudorascu DL, Vigen KK, Yu H, McKenzie CA, Sirlin CB, Brittain JH and Reeder SB. T(1) independent, T(2) (*) corrected chemical shift based fat-water separation with multi-peak fat spectral modeling is an accurate and precise measure of hepatic steatosis. J Magn Reson Imaging 2011; 33: 873-881.

[22] Meisamy S, Hines CD, Hamilton G, Sirlin CB, McKenzie CA, Yu H, Brittain JH, Reeder SB. Quantification of hepatic steatosis with T1-independent, T2-corrected MR imaging with spectral modeling of fat: blinded comparison with MR spectroscopy. Radiology 2011; 258: 767-75.

[23] Yokoo T, Shiehmorteza M, Hamilton G, Wolfson T, Schroeder ME, Middleton MS, Bydder M, Gamst AC, Kono Y, Kuo A, Patton HM, Horgan S, Lavine JE, Schwimmer JB, Sirlin CB. Estimation of hepatic proton-density fat fraction by using MR imaging at 3.0 T. Radiology 2011; 258: 749-59

[24] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011; 155: 529-36.

[25] Rinella ME, McCarthy R, Thakrar K, Finn JP, Rao SM, Koffron AJ, Abecassis M, Blei AT. Dual-echo, chemical shift gradient-echo magnetic

resonance imaging to quantify hepatic steatosis: implications for living liver donation. Liver Transpl 2003; 9: 851-6.
[26] Yoshimitsu K, Kuroda Y, Nakamuta M, Taketomi A, Irie H, Tajima T. Noninvasive estimation of hepatic steatosis using plain CT vs. chemical-shift MR imaging: significance for living donors. J Magn Reson Imaging 2008; 28: 678-84.
[27] d'Assignies G, Ruel M, Khiat A, Lepanto L, Chagnon M, Kauffmann C, Tang A, Gaboury L and Boulanger Y. Noninvasive quantitation of human liver steatosis using magnetic resonance and bioassay methods. Eur Radiol 2009; 19: 2033-2040.
[28] Kang BK, Yu ES, Lee SS, Lee Y, Kim N, Sirlin CB, Cho EY, Yeom SK, Byun JH, Park SH and Lee MG. A prospective comparison of magnetic resonance spectroscopy and analysis methods for chemical-shift gradient echo magnetic resonance imaging with histologic assessment as the reference standard. Invest Radiol 2012; 47: 368-375.
[29] Henninger B, Kremser C, Rauch S, Eder R, Judmaier W, Zoller H, Michaely H and Schocke M. Evaluation of liver fat in the presence of iron with MRI using T2* correction: a clinical approach. Eur Radiol 2013; 23: 1643-1649.
[30] Idilman IS, Keskin O, Celik A, Savas B, Halil Elhan A, Idilman R and Karcaaltincaba M. 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 2016; 57: 271-278.
[31] Imajo K, Kessoku T, Honda Y, Tomeno W, Ogawa Y, Mawatari H, Fujita K, Yoneda M, Taguri M, Hyogo H, Sumida Y, Ono M, Eguchi Y, Inoue T, Yamanaka T, Wada K, Saito S and Nakajima A. Magnetic resonance imaging more accurately classifies steatosis and fibrosis in patients with nonalcoholic fatty liver disease than transient elastography. Gastroenterology 2016; 150: 626-637.
[32] Graif M, Yanuka M, Baraz M, Blank A, Moshkovitz M, Kessler A, Gilat T, Weiss J, Walach E, Amazeen P, Irving CS. Quantitative estimation of attenuation in ultrasound video images: correlation with histology in diffuse liver disease. Invest Radiol 2000; 35: 319-24.
[33] Mottin CC, Moretto M, Padoin AV, Swarowsky AM, Toneto MG, Glock L, Repetto G. The role of ultrasound in the diagnosis of hepatic steatosis in morbidly obese patients. Obes Surg 2004; 14: 635-7.
[34] Limanond P, Raman SS, Lassman C, Sayre J, Ghobrial RM, Busuttil RW, Saab S, Lu DS. Macrovesicular hepatic steatosis in living related liver donors: correlation between CT and histologic findings. Radiology 2004; 230: 276-80.
[35] Fazel R, Krumholz HM, Wang Y, Ross JS, Chen J, Ting HH, Shah ND, Nasir K, Einstein AJ, Nallamothu BK. Exposure to low-dose ionizing radiation from medical imaging procedures. N Engl J Med 2009; 361: 849-57.
[36] Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. J Clin Epidemiol 1995; 48: 119-30.
[37] Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mosteller F. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med 1994; 120: 667-76.
[38] Walter SD. Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. Stat Med 2002; 21: 1237-56.
[39] Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The evidence-based medicine working group. JAMA 1994; 271: 703-7.

Table S1. Details of search strategy

	EMBASE	MEDLINE	COCHRANE	PUBMED
Patient	(exp fatty liver/or exp nonalcoholic fatty liver/or NAFLD.mp.) or (NASH.mp.) or (exp steatosis/) or (exp fat content/)	NAFLD OR NASH OR non-alcoholic fatty liver disease OR non-alcoholic steatohepatitis OR liver fat OR steatosis OR hepatic fat fraction OR liver fat content OR steatohepatitis OR hepatosteatosis	(NAFLD.mp. or exp Fatty Liver/) or (non-alcoholic fatty liver disease.mp.) or (non-alcoholic steatohepatitis.mp.) or (liver fat content.mp.)	(("Non-alcoholic Fatty Liver Disease" [Mesh]) OR "Fatty Liver" [Mesh]) and Humans [Mesh]
Intervention	(exp nuclear magnetic resonance imaging/) or (proton density fat fraction.mp. or proton/)	(magnetic resonance imaging or mri) OR proton density fat fraction	(exp Magnetic Resonance Imaging/) or (proton density fat fraction.mp.)	"Magnetic Resonance Imaging" [Mesh] and Humans [Mesh]
Number of articles retrieved	2932	1004	28	535

We used Medical Subject Headings (MeSH) terms and free text keywords: NAFLD, non-alcoholic fatty liver disease, liver or hepatic fat fraction, liver or hepatic fat content, steatosis, steatohepatitis, hepatosteatosis, NASH, non-alcoholic steatohepatitis. These search terms were combined with: magnetic resonance imaging, MRI, proton density fat fraction, PDFF.

Table S2. MR imaging features

Study	n of patients	Field strength (T)	Sequence	TR (ms)/TE (ms)	Breath holding	T2* correction	Evaluation of hepatic steatosis
Rinella 2003	22	1.5	T1-weighted dual gradient echo	180/2.38-4.76	NR	NR	IP-OP/IP + OP + OP/1.54
Yoshimitsu 2008	50	1.5	T1-weighted dual gradient echo	150/2.3-4.7	NR	NR	(SI (IP)-SI (OP))/SI (IP)
d'Assignies 2009	20	1.5	T1-weighted spoiled gradient dual echo	140/2.2-4.6	YES	YES	(SI (IP)-SI (OP))/2SI (IP)
Kang 2012	56	1.5	2-dimensional spoiled gradient echo, Multiecho and multi- interference	122/2.3-13.8	YES	YES	Sf×100/(Sf+Sw) #1
Benjamin 2013	31	1.5	two-dimensional transverse spoiled gradient-echo T1 fast low angle shot	103/2.37-5.05	YES	YES	(IP-OP)/(2×IP)
Tang,2013	77	1.5/3.0	two-dimensional spoiled gradient-recalled echo	1.5T:120-270/2.3-13.8 3.0:120-270/1.15-6.9	YES	YES	NR
Idilman 2013	70	1.5	T1-independent volumetric multiecho gradient-echo, including IDEAL-IQ	12.9/1.6-9.8	YES	YES	NR
Tang 2015	89	3.0	two-dimensional spoiled, gradient-recalled-echo	120-270/1.15-6.9	YES	YES	NR
Paparo,2015	77	1.5	two-dimensional, spoiled, multiecho gradient-echo	120-270/1.1-18.35	YES	YES	S2/(S1+S2) #2
Idilman 2016	19	1.5	T1-independent volumetric multi-echo gradient-echo, including IDEAL-IQ	12.9/1.6-9.8	YES	YES	NR
Imajo 2016	142	3.0	fast gradient echo (IDEAL-IQ)	110/2.1	YES	YES	NR

Abbreviations are as follows: n, number; TR, repetition time; TE, echo time; FS: Fat suppression; IP: In-phase; OP: Opposed-phase; SI: Signal intensity; NR, no report; IDEAL-IQ, iterative decomposition of water and fat with echo asymmetry and least squares estimation-image quantification. #1: Sw = SI of the water component, Sf = SI of the fat component. #2: S1 and S2 are the signal amplitudes of water and fat respectively.

Table S3. MRI-PDFF diagnosis accuracy

Group	Steatosis grade	Study	TP	FP	FN	TN	Sensitivity (%)	Specificity (%)	ROC (SE)	Cut-off
Grade 0	0-5%	Rinella, 2003	16	0	2	4	88.9	100	NR	>15%,0.2,
		d'Assignies, 2009	20	3	1	21	95.2	87.5	0.99	0.05
		Kang, 2012 Multiecho	17	7	1	31	94.4	81.6	0.947	2.51
		Kang, 2012 Multi-Interference	17	7	1	31	94.4	81.6	0.948	2.94
		Benjamin, 2013	20	2	3	6	87	75	0.84	0.053
		Tang, 2013	70	0	2	5	97	100	0.989	0.064
		Tang, 2015	71	1	12	5	86	83	0.961	0.069
		Paparo, 2015	27	1	4	45	87.10	97.83	0.926	0.0687
		Imajo, 2016	-	-	-	-	90.0	93.3	0.98	0.052
		Summary estimates (± 95% CI)					0.91 (0.87,0.94)	0.88 (0.82,0.92)	0.9594 (0.0139)	-
Grade 1	5%-33%	Yoshimitsu, 2008	14	1	2	17	87.5	94.4	0.91	0.03
		Summary estimates (± 95% CI)					-	-	-	-
Grade 2	33-66%	Yoshimitsu, 2008	25	4	0	5	100	55.6	0.88	0.25
		Tang, 2013	28	3	18	28	61	90	0.825	0.174
		Idilman, 2013	41	4	3	22	93	85	0.95	0.1503
		Tang, 2015	28	2	16	43	64	96	0.947	0.164
		Paparo, 2015	7	8	1	61	87.5	88.41	0.929	0.1108
		Idilman, 2016	12	2	0	5	100	71.4	0.881	0.1
		Imajo, 2016	-	-	-	-	78.9	84.1	0.90	0.113
		Summary estimates (± 95% CI)					0.79 (0.72,0.85)	0.88 (0.82,0.92)	0.9298 (0.0195)	-
Grade 3	>66%	Tang, 2013	13	5	6	53	68	91	0.893	0.221
		Tang, 2015	10	6	4	69	71	92	0.921	0.235
		Imajo, 2016	-	-	-	-	73.7	81.0	0.79	0.171
		Summary estimates (± 95% CI)					-	-	-	-

Abbreviations are as follows: TP, true positive; FP, false positive; FN, false negative; TN, true negative; ROC, the area under receiver operating characteristic curves; SE, standard error.