Abdominal Radiology

Published online: 27 September 2016

Abdom Radiol (2017) 42:490–501
DOI: 10.1007/s00261-016-0913-6

Liver fibrosis staging with diffusion-weighted imaging: a systematic review and meta-analysis

Hanyu Jiang, Jie Chen, Ronghui Gao, Zixing Huang, Mingpeng Wu, Bin Song

Department of Radiology, Sichuan University West China Hospital, No. 37 Guoxue Alley, Chengdu, Sichuan, China

Abstract

Purpose: A meta-analysis was performed to assess the diagnostic performance of diffusion-weighted imaging (DWI) in liver fibrosis (LF) staging.

Methods: We conducted a comprehensive literature search to identify relevant articles. Diagnostic data were extracted for each METAVIR fibrosis stage (F0–F4). A bivariate binomial model was used to combine sensitivities and specificities. Summary receiver operating characteristics (SROC) curves were performed and areas under SROC curve (AUC) were calculated to indicate diagnostic accuracies. Subgroup analyses were performed between different study characteristics.

Results: Twelve studies met the inclusion criteria for LF ≥F1, 16 for ≥F2, 18 for ≥F3, and 12 for F4. AUCs of DWI were 0.8554, 0.8770, 0.8836, and 0.8596 for ≥F1, ≥F2, ≥F3, and F4, respectively. Subgroup analyses showed that for LF ≥F2 and ≥F3, maximal *b* values (b_{max}) ≥ 800 s/mm^2 performed significantly better than b_{max} < 800 s/mm^2. The diagnostic accuracies of 3.0 T and intravoxel incoherent motion (IVIM)-DWI were significantly higher than those of 1.5 T and conventional DWI for diagnosing liver cirrhosis (F4).

Conclusions: DWI is a reliable noninvasive technique with good diagnostic accuracy for LF staging. Using b_{max} ≥ 800 s/mm^2, high-field strength (3.0 T) and IVIM-DWI can optimize the diagnostic performance of DWI.

Key words: Diffusion magnetic resonance imaging—Liver cirrhosis—Sensitivity—Specificity—Meta-analysis

Abbreviations

ADC	Apparent diffusion coefficient
AUC	Area under SROC curve
b_{max}	Maximal *b* value
CI	Confidence interval
CLD	Chronic liver disease
DWI	Diffusion-weighted imaging
ECM	Extracellular matrix
FN	False-negative
FP	False-positive
HCC	Hepatocellular carcinoma
IVIM	Intravoxel incoherent motion
MR	Magnetic resonance
MRE	Magnetic resonance elastography
MRI	Magnetic resonance imaging
NLR	Negative likelihood ratio
PLR	Positive likelihood ratio
QUADAS-2	Quality assessment of diagnostic accuracy studies-2
SNR	Signal-to-noise ratio
SROC	Summary receiver operating characteristics
TN	True-negative
TP	True-positive

Liver fibrosis (LF) is the most frequent consequence of all chronic liver diseases (CLDs) [1], characterized by the excessive accumulation of extracellular matrix (ECM) [2], leading to the replacement of injured tissue by collagenous scar and the consequent liver architectural distortion. The major clinical consequences of cirrhosis are impaired liver function, portal hypertension, and the development of hepatocellular carcinoma (HCC) [3]. The end-stage LF is often considered irreversible with very limited effective treatment except liver transplantation, whereas early or intermediate hepatic fibrosis is usually a treatable complication [4, 5]. Therefore, early detection and staging of LF is crucial for therapeutic decision-making and monitoring treatment responses.

Hanyu Jiang and Jie Chen have contributed equally to this work.

Correspondence to: Bin Song; *email:* anicesong@vip.sina.com

Currently, biopsy is considered the gold standard for assessing LF [6]. However, it is invasive and prone to sampling variability [7]. Therefore, noninvasive assessments for the evaluation of LF have become a heated discussed topic worldwide [8–10]. Diffusion-weighted imaging (DWI) is a specific functional magnetic resonance imaging (MRI) technique based on the principles of Brownian motion (random thermal diffusion) of small molecules in a tissue [11]. As a notable DWI-based imaging technique, Intravoxel incoherent motion (IVIM) analyses the signal decay of multiple b values to simultaneously evaluate the perfusion-related diffusivity (demonstrated by parameters D^*, f) and pure molecular diffusivity (demonstrated by the parameter D) [12]. An increasing number of studies have been focused on the diagnostic performances of DWI for the staging of LF [13, 14], although discrepant results have been reported among those studies.

A previous study [15] compared the diagnostic accuracies of magnetic resonance elastography (MRE) and DWI for the assessment of LF, and concluded that MRE is more reliable for LF staging. In our opinion, with more studies and patients included in this meta-analysis, although the performance of DWI was limited, MRE is currently available only in selected centers while DWI is a widely available and easy-to-perform technique. Recent advances in DWI techniques have showed progresses for LF staging. This study aims to evaluate the diagnostic performance of DWI in LF staging and explore factors that may influence the diagnostic accuracy.

Materials and methods

Literature search and screening

A systematic literature search was performed by two investors independently in MEDLINE, Web of Science, EMBASE, Springer Link, and Science Direct to identify relevant articles published before February 2016 with the keywords "liver/hepatic fibrosis or cirrhosis" and "diffusion magnetic resonance imaging or diffusion-weighted imaging or DWI or apparent diffusion coefficient or intravoxel incoherent motion-DWI." The research was limited to articles concerning humans with an abstract in English.

Two reviewers read the titles and abstracts of the yielded articles which addressed the diagnostic performance of DWI for staging LF in humans to select potentially relevant articles. The full set of selected articles was collected and reviewed independently by the same reviewers to determine their eligibility for further quantitative analysis. The inclusion criteria were as follows: (1) DWI was performed to identify LF; (2) sufficient data were available to calculate true-positive (TP), false-positive (FP), false-negative (FN), and true-negative (TN) values; (3) histopathology (METAVIR score) as the reference standards; and (4) the study population should be no less than 20. The exclusion criteria were as follows: (1) duplicate publication based on the same primary study; (2) articles with poor quality; (3) studies focused on children; and (4) nonoriginal researches including review articles, abstracts, letters, comments, guidelines and case reports. Investigators were not blinded to the information about the authors, the authors' affiliation, or the journal name. Disagreements between the two reviewers were resolved by consensus. Investigators of the primary researches were approached for additional information, if necessary.

Data extraction and quality assessment

Data were extracted by the same two reviewers mentioned above independently. A senior radiologist with more than 20 years of experience in hepatic disease diagnosis was consulted to resolve discrepancy between the two reviewers.

To extract data concerning study characteristics, we recorded the patient information (study population, number of male and female patients, mean patient age with range and patient spectrum), study design (prospectively or retrospectively), score system for histopathologic staging LF, blinding procedure, reference standard (i.e., liver biopsy and/or surgery), and time interval between index test and reference standard. We also recorded the image protocols (magnetic field strength, b values, and MR scanner) adopted in the primary studies to perform DWI. For the calculation of diagnostic accuracy of DWI, we extracted available data on TPs, FNs, FPs, and TNs. We grouped accuracy results into five subgroups which were F0 = no fibrosis; F1 = portal fibrosis without septa; F2 = portal fibrosis and few septa; F3 = numerous septa without cirrhosis; and F4 = cirrhosis [16]. The 2 × 2 contingency tables were formed for the calculation of F0 vs. F1–F4 (≥F1), F0 and F1 vs. F2–F4 ((≥F2), F0–F2 vs. F3 and F4 ((≥F3), and F0–F3 vs. F4, respectively. The quality of the included studies was assessed according to quality assessment of diagnostic accuracy studies-2 (QUADAS-2) [17].

Statistical analysis

We first used a random-effects coefficient binary regression model to summarize the diagnostic performances. The summary receiver operating characteristic (SROC) curve was constructed and areas under the SROC curve (AUCs) of each LF stage served as the determination of the diagnostic performance of DWI [18].

Heterogeneity between the included studies was evaluated. Several potential sources may contribute to the heterogeneity, the first of which is the threshold effect. We confirmed the absence of threshold effect by not noticing the "shoulder-arm" shape in the SROC plane [19].

Apart from variations due to threshold effect, heterogeneity could be generated from other related factors. The heterogeneity was identified by the Q statistic of the χ^2 value test and the inconsistency index (I^2), and $p < 0.1$ or $I^2 > 50\%$ indicated the presence of heterogeneity [20]. If significant heterogeneity was detected, we then performed single-factor meta-regression analyses to determine factors that contributed to the heterogeneity and subgroup analyses to observe their quantitative effects on the diagnostic results [21]. The subgroup analyses of different LF stages included comparisons of (1) study design (prospective vs. retrospective); (2) blinding procedure (yes vs. unclear); (3) study population; (4) mean patient age; (5) patient gender (male vs. female); (6) reference standard (liver biopsy vs. surgery); (7) MR scanner; (8) MR field strength (1.5 T vs. 3.0 T); (9) number of b values; (10) maximal b value ($b_{max} \geq 800$ s/mm^2 vs. $b_{max} < 800$ s/mm^2); and (11) MR modality (conventional DWI vs. IVIM).

Publication biases were assessed with the Deek's funnel plots and Egger's asymmetry tests [22]. An inverted symmetrical funnel plot with $P > 0.05$ was considered to indicate the absence of publication bias [23].

Results

Study selection and quality assessment

The systematic search initially yielded 301 results, of which 25 studies were included in this meta-analysis. All studies were published between October, 2007 [24] and February, 2016 [25]. The study flowchart is demonstrated in Fig. 1. The qualities of included studies were good. Quality assessment for the included diagnostic studies is presented in Table 1. Figure 2 shows a graphical display for QUADAS-2 results regarding the proportion of studies with low, high, or unclear risk of bias.

Study characteristics

The 25 included studies [24–48] involved 1833 patients, whose ages ranged from 16 to 89. 14 trials [24, 26–28, 30–34, 38, 41, 44, 45, 47] were prospective, with the remaining 11 studies [25, 29, 35–37, 39, 40, 42, 43, 46, 48] retrospective. Blinding procedure was reported in 15 studies [24, 25, 28–30, 34–39, 43, 44,

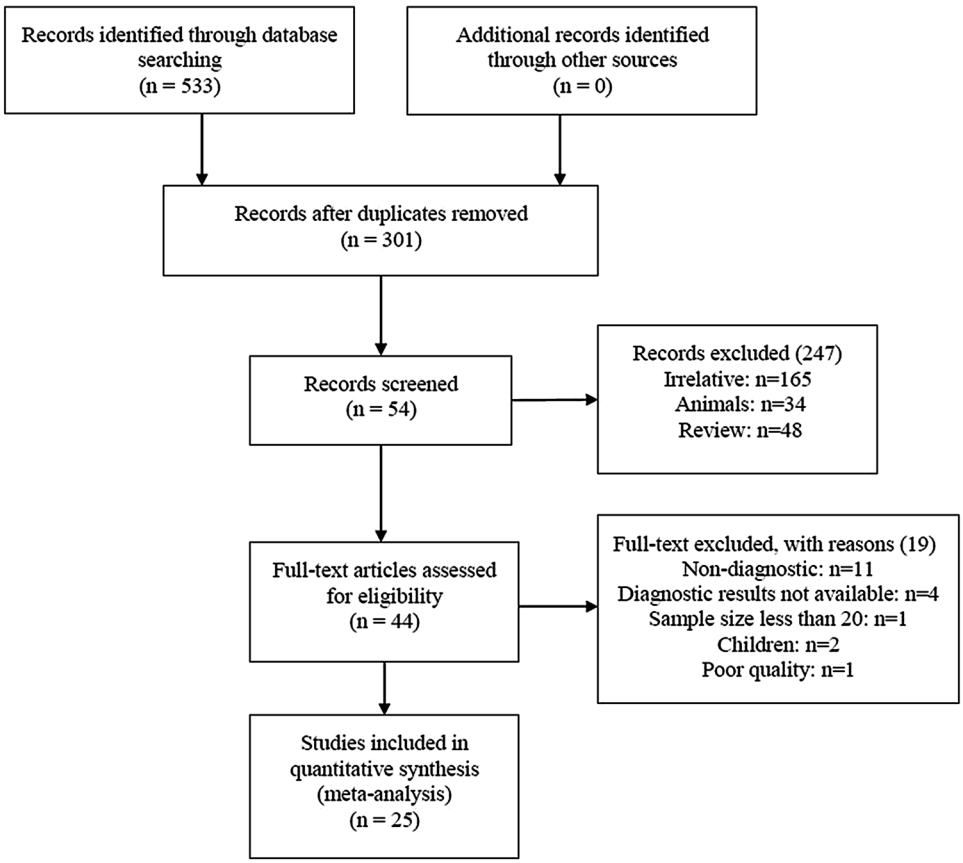

Fig. 1. Study flowchart.

Table 1. Quality assessment of the 25 included diagnostic studies

Study	RISK OF BIAS				APPLICABILITY CONCERNS		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Parente et al	?	☺	☺	☺	☺	☺	☺
Ichikawa et al	☺	☺	☺	?	☺	☺	☺
Kocakoc et al	☹	?	☺	☺	☺	☺	☺
Wu et al	☺	☺	☺	☺	☺	☺	☺
Feier et al	☺	☺	☺	☺	☺	☺	☺
Ding et al	☺	☺	☺	☺	☺	☺	☺
Hong et al	☺	☺	☺	?	☺	☺	☺
Chen et al	☺	?	☺	☺	☺	☺	☺
Yoon et al	☺	?	☺	☹	☺	☺	☺
Bonekamp et al	☺	?	☺	☺	☺	☺	☺
Tokgoz et al	☺	☺	☺	?	☺	☺	☺
Chung et al	☺	☺	☺	☺	☺	☺	☺
Telena et al	☺	☺	☺	☺	☺	☺	☺
Vaziri-Bozorg et al	?	?	☺	☺	☺	☺	☺
Bonekamp et al	☺	☺	☺	☺	☺	☺	☺
Wang et al	☺	☺	☺	☺	☺	☺	☺
Fujimoto et al	☹	☺	☺	?	☺	☺	☺
Richard et al	☺	?	☺	?	☺	☺	☺
Patel et al	☺	?	☺	☹	☺	☺	☺
Shi et al	☹	?	☺	☺	☺	☺	☺
Sandrasegaran et al	☺	☺	☺	?	☺	☺	☺
Zhou et al	☹	?	☺	☺	☺	☺	☺
Taouli et al	☺	?	☺	☺	☺	☺	☺
Taouli et al	☺	☺	☺	?	☺	☺	☺
Lewin et al	?	☺	☺	☺	☺	☺	☺

☺, low risk; ☹, high risk; ?, unclear risk

46, 47] and the rest 10 unclear [26, 27, 31–33, 40–42, 45, 48]. The disease spectrum was restricted to chronic hepatitis in six trials [27, 29–33], type 2 diabetic patients in one trial [28], and in the remaining 18 trials, there was no restriction (Table 2). The parameters of imaging acquisition was demonstrated on Table 3.

For study-level analyses, 12 studies [25, 26, 28, 29, 33–35, 38, 41–44] met the inclusion criterial for fibrosis stage ≥F1, 16 studies [24–27, 29, 31, 34, 35, 37, 40, 42–44, 47, 48] ≥F2, 18 studies [24–27, 29–31, 34–39, 42–44, 47, 48] ≥F3, and 12 studies [25, 29, 32, 34, 35, 37, 40, 42–45, 48] F4. All patients had biopsy or surgery results as reference standards. Diagnostic results of each subset are presented in Table 4.

Diagnostic performance

Pooled sensitivities with corresponding 95% confidence intervals (CIs) for LF ≥F1, ≥F2, ≥F3, and F4 were 0.78 (95% CI 0.75–0.82), 0.81 (95% CI 0.78–0.84), 0.71 (95% CI 0.67–0.75), and 0.80 (95% CI 0.75–0.85), respectively. Pooled specificities for LF ≥F1, ≥F2, ≥F3, and F4 were 0.78 (95%CI 0.73–0.82), 0.80 (95% CI 0.76–0.83), 0.84 (95% CI 0.81–0.86), and 0.77 (95% CI 0.74–0.81), respectively. According to the SROC curve, the AUCs of LF ≥F1, ≥F2, ≥F3, and F4 were 0.8554, 0.8770, 0.8836, and 0.8596, respectively. Forest plots of sensitivity, specificity, positive likelihood ratio (PLR), and negative likelihood ratio (NLR) of different subgroups are shown in Table 4. The SROC curves are shown in Fig. 3.

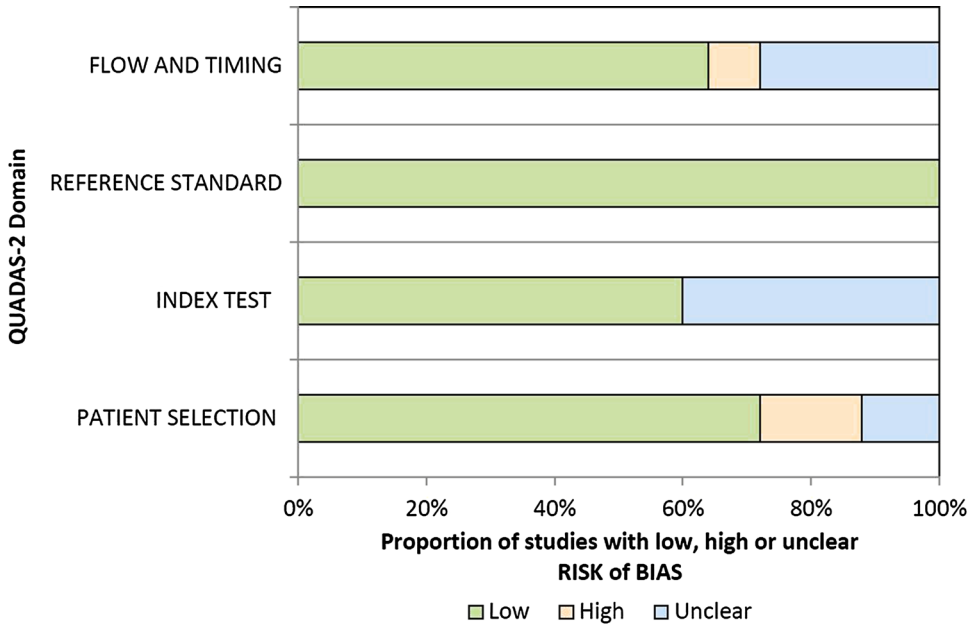

Fig. 2. Graphical display for QUADAS-2 results regarding proportion of studies with low, high, or unclear risk of bias.

Heterogeneity assessing and meta-regression analysis

Highly significant heterogeneity was detected in this meta-analysis. Threshold effects of all the fibrosis stages were eliminated through the SROC planes, which showed no "shoulder-arm" shapes. The single-factor meta-regression analyses showed that for LF ≥F1, no factor contributed statistically significantly to heterogeneity; for both LF ≥F2 and ≥F3, the maximal b values were the most important variable source of heterogeneity; while magnetic field strength and the MR imaging protocols contributed mostly to the heterogeneity of F4. Study design, patient age, patient gender, reference standard, numbers of b values, and blinding procedure did not contribute statistically to the heterogeneity in any fibrosis stage.

Subgroup analyses

We performed subgroup analyses between different study characteristics in each fibrosis stage to evaluate their quantitative effects on heterogeneity. The sensitivity, specificity, and AUC of $b_{max} < 800$ s/mm^2 for LF ≥F2 were 0.75, 0.71, and 0.7994, respectively, and those of $b_{max} \geq 800$ s/mm^2 were 0.85, 0.86, and 0.9183, respectively, and difference was statistically significant. The sensitivity, specificity, and AUC of $b_{max} < 800$ s/mm^2 for LF ≥F3 were 0.59, 0.83, and 0.8360, respectively, and those of $b_{max} \geq 800$ s/mm^2 were 0.82, 0.84, and 0.9162, respectively, demonstrating a statistically significant difference. For LF = F4, statistically significant differences were detected between the diagnostic accuracies of different magnetic field strengths ($p = 0.0354$) and the MR modalities ($p = 0.0335$). The results of the subgroup analyses are presented in Table 5.

Publication biases

The funnel plot shows that studies were distributed symmetrically on a scatter plot. The p values of the Deeks' funnel plot asymmetry test for LF ≥F1, ≥F2, ≥F3, and F4 were 0.35, 0.37, 0.82, and 0.29, respectively, which demonstrated no evidence of notable publication bias (Fig. 4).

Discussion

DWI is a quick and repeatable noninvasive MR modality which enables qualitative and quantitative evaluation of tissue diffusivity without the use of gadolinium chelates. In LF staging, the apparent diffusion coefficient (ADC) of the fibrotic hepatic tissue is usually significantly lower than that of normal liver tissues, and the ADC values decrease as the fibrosis score increases [49, 50]. A possible explanation for this phenomenon suggested that in fibrotic liver tissues, with the presence of increased proton poor connective tissue, the molecular diffusion and the blood flow were restricted [11, 12, 51], leading to the decreased ADC in these tissues. Previous studies have shed light on the feasibility of DWI in the staging of LF, monitoring treatment responses and follow-up of patients with LF [24–48].

In this meta-analysis, we first explored the ability of DWI in LF staging. A diagnostic tool is defined as per-

Table 2. Study and patient characteristics of included studies

Study and date	Design	Blind	No. patients	Age	M/F	F0/F1/F2/F3/F4	Disease spectrum
Parente et al. 2015	Prospective	Yes	59	54	49/10	43/9/5/2[&]	Type 2 diabetic patients
Ichikawa et al. 2015	Retrospective	Yes	182(129/53)	66.4(19–86)	127/55	NA	HCV, HBV, ASH, NASH, AIH, et al
Kocakoc et al. 2015	Prospective	Unclear	74(44/30)	31–60	28/16	NA	HBV, HCV
Wu et al. 2015	Prospective	Yes	49	62.7(38–85)	36/13	6/16/10/10/7	HCC, metastasis, CCC, et al
Feier et al. 2016	Retrospective	Yes	77	58(18–81)	47/30	21/7/8/12/29	HCV, HBV, NASH, ASH, PSC, et al
Ding et al. 2015	Retrospective	Yes	145	54(24–75)	115/30	34/7/12/14/78	HCC, intrahepatic CCC, cirrhosis, FNH, et al.
Hong et al. 2014	Retrospective	Yes	76	42.3	43/33	18/14/14/20/10	CLD, focal hepatic lesions
Chen et al. 2014	Prospective	Unclear	50(25/25)	43.7(25–73)	22/3	25/4/9/11	HBV, HCV, alcoholism, unknown
Yoon et al. 2014	Retrospective	Unclear	55	53.9(18–78)	42/13	11/7/7/9/21	HBV, HCV, alcohol induced
Bonekamp et al. 2014	Retrospective	Unclear	85	48.6	50/35	35/17/2/724	HBV, HCV, HDV, cholangitis, biliary obstruction, ALD, et al.
Tokgoz et al. 2014	Prospective	Yes	138(68/70)	59	NA	70/32/15/7/14	HBV, HCV, NASH, ASH
Chung et al. 2015	Retrospective	Yes	57	58.7(32–89)	35/22	21/1/6/7/22	HBV, HCV, ALD, HCC, CCC, metastasis
Kovač et al. 2012	Prospective	Yes	45	48	9/36	12/3/6/11/13	PBC, PSC
Vaziri-Bozorg et al. 2012	Prospective	Unclear	44(33/11)	37.93	29/4	NA	HBV,HCV
Bonekamp et al. 2011	Retrospective	Yes	88	50(44–55)	53/35	33/20/2/6/27	HCV, HBV, ALD, NAFLD, cholestatic liver disease, et al.
Wang et al. 2011	Prospective	Yes	76	55(20–74)	50/26	32/12/6/6/20	HCV, HBV, NASH, AIH, et al.
Fujimoto et al. 2011	Retrospective	Yes	55(43/12)	65(34–83)	35/20	12/9/11/11/12	HCV
Do et al. 2010	Retrospective	Unclear	56(34/22)	57(18–72)	26/8	22/4/5/4/21	HCV, HBV, PBC, ALD, NAFLD et al.
Patel et al. 2010	Prospective	Unclear	30	48(23–89)	18/12	(Non-:cirrhotic) 16:14	HCV, alcoholism, AIH
Shi et al. 2010	Prospective	Unclear	59(47/12)	38(19–57)	30/17	20/6/9/10/14	HBV, HCV
Sandrasegaran et al. 2009	Retrospective	Yes	78	53(28–74)	55/23	11/16/10/14/27	HCV, HBV, NASH, alcoholism, AIH, et al.
Zhou et al. 2009	Prospective	Unclear	107(85/22)	33(16–57)	53/32	34/26/22/15/10	Chronic hepatitis
Taouli et al. 2008	Prospective	Unclear	44(31/13)	47.4	24/7	15/7/2/6/14	HCV, HBV, alcoholism, liver steatosis
Taouli et al. 2007	Prospective	Yes	30(23/7)	54(39–77)	15/8	11/5/4/4/6	HCV, HBV, NASH, AIH, alcoholism
Lewin et al. 2007	Prospective	Yes	74(54/20)	46.4(25–75)	37/17	1/30/8/5/10	HCV

M, male; F, female; NA, data not available; AIH, autoimmune hepatitis; ALD, alcoholic liver disease; ASH, alcoholic steatohepatitis; CCC, cholangiocarcinoma; CLD, chronic liver disease; FNH, focal nodular hyperplasia; HBV, hepatitis B virus; HCV, hepatitis C virus; HDV, hepatitis D virus; HCC, hepatocellular carcinoma; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; PBC, primary biliary cirrhosis; PSC, primary sclerosing cholangitis

Table 3. DWI imaging protocols

Study	Reference test	Histology study model	Time intervals	MRI scanners	FS	No. b	*b* values
Parente et al.	Biopsy	Metavir	≤3 M	Philips Achieva MR system	3.0 T	10	0,10,20,40,80,160,200,400,800,1000
Ichikawa et al.	Biopsy/surgery	Metavir	≤3 M	Discovery 750, GE Medical Systems	3.0 T	11	0,10,20,30,40,50,80,100,200,500,1000
Kocakoc et al.	Biopsy	Modified Ishak	≤1 D	GE Healthcare	1.5 T	3	100,600,1000
Wu et al.	Surgery	Metavir	≤7 D	Magnetom Verio, Siemens Medical Solutions	3.0 T	16	0,10,20,30,40,50,60,70,80,90,100,200, 300,400,500,1000
Feier et al.	Biopsy	Metavir	1.6±2.0M(0–6)	Magnetom Trio, Siemens Healthcare	3.0 T	3	50,300,600
Ding et al.	Surgery	Metavir	3 D(2–7)	Magnetom Aera, Siemens medical solution	1.5 T	2	0,500
Hong et al.	Surgery/biopsy	Metavir	27 D(14–60)	Signa HDxt, GE Healthcare	3.0 T	3	200,600,800
Chen et al.	Biopsy	Metavir	≤30 D	GE Medical system	3.0 T	7	0,50,100,200,400,600,800
Yoon et al.	Surgery/biopsy	Metavir	Unclear	Magnetom Verio, Siemens Healthcare	3.0 T	8	0,25,50,75,100,200,500,800
Bonekamp et al.	Biopsy	Metavir	≤12 M	GE Signa, GE medical system/Siemens Magnetom Avanto, Siemens Healthcare	1.5 T	2	0,750
Tokgoz et al.	Biopsy	Metavir	Unclear	Intera, Master Gyroscan, Philips Medical system	1.5 T	2	0,600
Chung et al.	Surgery	Metavir	15.9 D(2–43)	Magnetom Avanto, Siemens Medical Solution	1.5 T	9	0,30,60,100,150,200,400,600,900
Kovač et al.	Biopsy	Metavir	2 M15D-5M	Magnetom Avanto, Siemens Medical Systems	1.5 T	5	0,50,200,400,800
Vaziri-Bozorg et al.	Biopsy	Ishak	≤1 M	GE Signa, GE Medical Systems	1.5 T	2	0,500,700,1000
Bonekamp et al.	Biopsy	Metavir	73 D(0–322)	GE Signa/Siemens Magnetom Avanto	1.5 T	2	0,750
Wang et al.	Biopsy	Metavir	60 D(1–359)	Magnetom Espree, Siemens Medical Solution	1.5 T	3	50,500,1000
Fujimoto et al.	Surgery/biopsy	Metavir	7 D(1–14)	Magnetom Symphony Advanced, Siemens	1.5 T	2	0,1000
Do et al.	Surgery/biopsy	Batts-Ludwing	32 D(0–85)	Magnetom Avanto, Siemens Healthcare	1.5 T	3	0,50,500
Patel et al.	Biopsy/MRI	Unclear	Unclear	Magnetom Avanto, Siemens	1.5 T	9	0,50,100,150,200,300,500,700,1000
Shi et al.	Biopsy	Metavir	31 D(20–43)	Gyroscan intera, Philips	1.5 T	5	0,250,500,750,1000
Sandrasegaran et al.	Surgery/biopsy	Metavir	2.4 M(0–6)	Magnetom Avanto, Siemens Healthcare	1.5 T	2	50,400
Zhou et al.	Biopsy	Metavir	1 D	Magnetom Avanto, Siemens	1.5 T	5	100,300,500,800,1000
Taouli et al.	Biopsy	Batts-Ludwig	52 D(0–372)	Gyroscan and Intera systems, Philips Medical Systems	1.5 T	2	0,500
Taouli et al.	Biopsy/MRI	Batts-Ludwig	42 D(9–70)	Magnetom Avanto, Siemens Medical Solutions	1.5 T	2	0,700
Lewin et al.	Biopsy	Metavir	0 D	Magnetom Maestro Class Symphony, Siemens Medical	1.5 T	4	0,200,400,800

FS, field strength; D, days; M, months

Table 4. Diagnostic results of different LF stages

	No. studies	Sensitivity	Specificity	PLR	NLR	AUC
≥F1	12	0.78 (0.75–0.82)	0.78 (0.73–0.82)	3.24 (2.43–4.32)	0.29 (0.22–0.37)	0.8554
≥F2	16	0.81 (0.78–0.84)	0.80 (0.76–0.83)	3.72 (2.75–5.02)	0.25 (0.19–0.33)	0.8770
≥F3	18	0.71 (0.67–0.75)	0.84 (0.81–0.86)	4.44 (3.35–5.88)	0.28 (0.18–0.42)	0.8836
≥F4	12	0.80 (0.75–0.85)	0.77 (0.74–0.81)	3.30 (2.66–4.09)	0.31 (0.23–0.41)	0.8596

LF, liver fibrosis; PLR, positive likelihood ratio; NLR, negative likelihood ratio; AUC, area under SROC curve

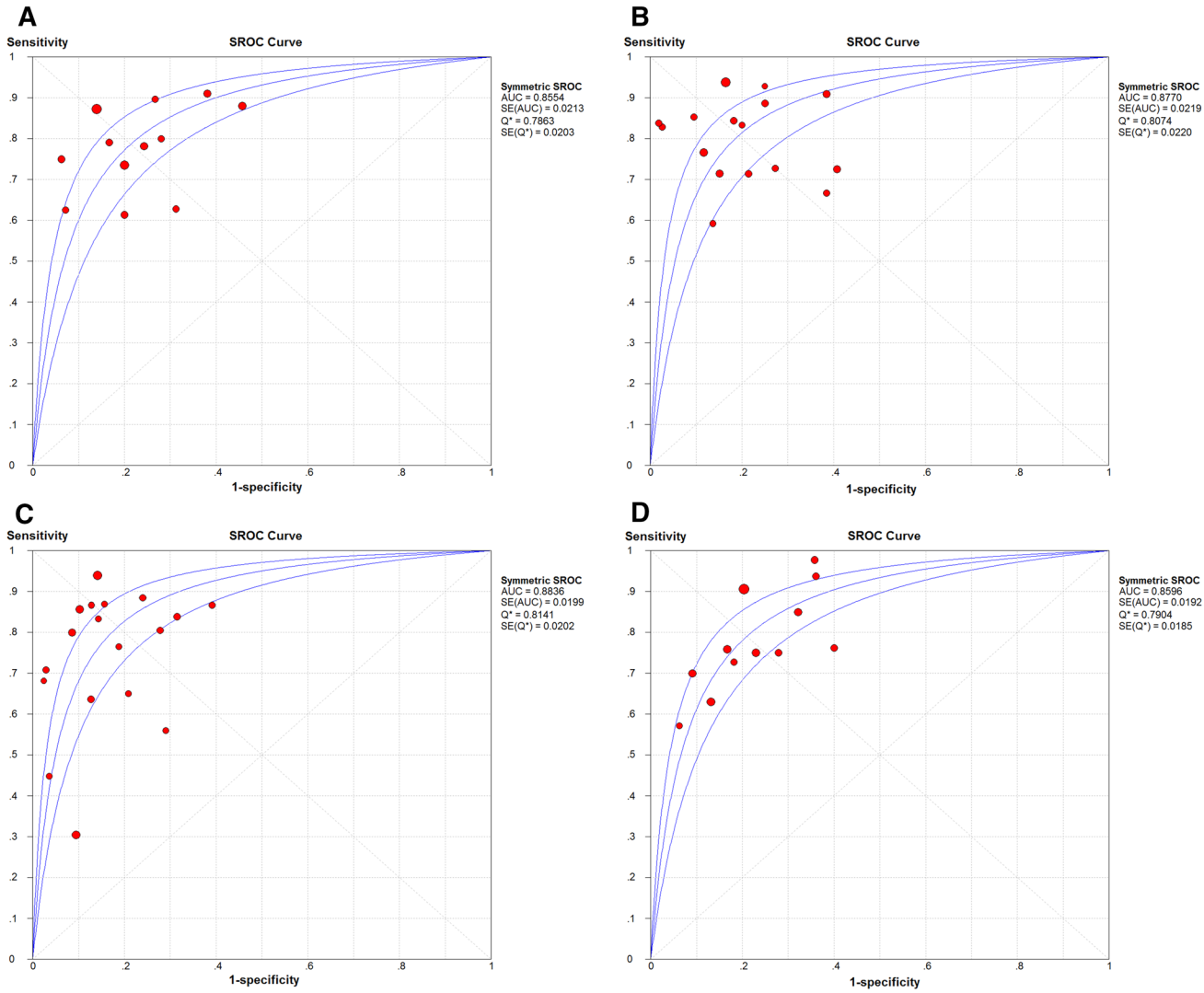

Fig. 3. Summary receiver operating characteristic (SROC) curve for DWI in staging LF. The AUCs of LF ≥F1 (**A**), ≥F2 (**B**), ≥F3 (**C**), and ≥F4 (**D**) were 0.8554, 0.8770, 0.8836, and 0.8596, respectively, indicating a good but not excellent diagnostic accuracy.

fect if the AUC is 100%, excellent if the AUC is greater than 90%, and good if the AUC is greater than 80% [52]. According to this, the results showed that DWI had good but not excellent diagnostic accuracy for LF staging.

To our knowledge, there have not been any standardized DWI techniques as yet, and a large variety of imaging parameters exist for DWI in the number and size of b values, and diagnostic threshold for different tissues, organs, and diseases. At least two b factors are required for the calculation of ADC when performing DWI. Although several b values are often used in practice to perform a linear regression analysis to lower perfusion contamination and regional ADC variation to consolidate the ADC evaluation [53], our meta-regression analyses revealed that the number of b values did not

Table 5. Integrant results of the subgroup analyses

	No. study	Sensitivity	Specificity	PLR	NLR	AUC	*p* value[a]
≥F2							
All	16	0.81(0.78–0.84)	0.80(0.76–0.83)	3.72(2.75–5.02)	0.25(0.19–0.33)	0.8770	
Maximal *b* value							<0.01
b_{max} < 800	7	0.75(0.69–0.81)	0.71(0.65–0.77)	2.52(1.93–3.30)	0.37(0.28–0.49)	0.7994	
b_{max} ≥ 800	9	0.85(0.80–0.88)	0.86(0.82–0.90)	5.22(3.96–6.87)	0.19(0.13–0.29)	0.9183	
≥F3							
All	18	0.71(0.67–0.75)	0.84(0.81–0.86)	4.44(3.35–5.88)	0.28(0.18–0.42)	0.8836	
Maximal *b* value							<0.01
b_{max} < 800	8	0.59(0.53–0.65)	0.83(0.79–0.87)	3.63(2.44–5.38)	0.39(0.23–0.64)	0.8360	
b_{max} ≥ 800	10	0.82(0.77–0.86)	0.84(0.81–0.88)	5.37(3.53–8.17)	0.21(0.12–0.37)	0.9162	
F4							
All	12	0.80(0.75–0.85)	0.77(0.74–0.81)	3.30(2.66–4.09)	0.31(0.23–0.41)	0.8596	
Modality							0.03
ADC	8	0.74(0.66–0.81)	0.79(0.74–0.83)	3.31(2.49–4.39)	0.35(0.26–0.45)	0.8257	
*D**	4	0.89(0.81–0.94)	0.75(0.69–0.81)	3.34(2.27–4.91)	0.16(0.04–0.59)	0.9155	
Field strength							0.04
1.5 T	7	0.72(0.64–0.80)	0.76(0.71–0.81)	2.92(2.25–3.80)	0.38(0.29–0.50)	0.8166	
3.0 T	5	0.88(0.81–0.93)	0.79(0.74–0.83)	3.74(2.65–5.28)	0.19(0.10–0.37)	0.9023	

[a] Represents the *p* value of meta-regression analysis, $p \leq 0.05$ indicates significant contribution to heterogeneity
PLR, positive likelihood ratio; NLR, negative likelihood ratio; AUC, area under SROC curve

statistically significantly correlate with the diagnostic performances of DWI in the staging of LF.

Typical b values for LF imaging vary from 0 to 1000 s/mm^2. In subgroup analysis, our study revealed that for fibrosis stage ≥F2 and ≥F3, $b_{max} \geq 800$ s/mm^2 performed statistically significantly better compared to $b_{max} < 800$ s/mm^2 in the staging of LF, indicating that adopting b_{max} between 800 and 1000 s/mm^2 could significantly optimize the diagnostic accuracy of DWI in the staging of LF for significant and severe fibrosis (F2 and greater). In clinical practices, high accuracies in the detection of ≥F2 and ≥F3 are essential. Owing to cost, risk of toxicity, and limited efficacy, Kim et al. [54] suggested that only hepatitis C patients whose LF ≥F2 should receive antiviral treatment. Moreover, significant fibrosis (F2) is usually considered as a hallmark of a progressive disease, and the major treatment for this fibrosis stage is resolving the underlying cause of liver disease [15, 55]. Apart from these, discrimination of advanced fibrosis (F3) or cirrhosis (F4) is essential because those patients should be screened for portal hypertension and HCC [55].

This finding was in accordance with previous studies. Ozkurt et al. [56] used different b values including 250, 500, 750, and 1000 s/mm^2 in their study and found that the negative correlation between the fibrosis score and ADC values were significant only in b values of 750 and 1000 s/mm^2. Taouli et al. [24] reported that the ADC value was significantly correlated with the LF stage with b values of at least 500 s/mm^2, and the ADC value with the highest significant correlation with fibrosis stage was acquired from a combination of b values of 0–1000 s/mm^2. Other studies which involved low b values (e.g., 0–128 s/mm^2 and 50–400 s/mm^2) reported that significant correlations with the ADC values and hepatic fibrosis stage were not achieved [30, 38, 46, 53]. A possible explanation for this is that with low b values, blood flow will contribute more to the signal attenuation. Therefore, relatively small b_{max} could increase the amount of perfusion contamination in ADC measurement [11]. However, we did not identify statistically significant differences between the diagnostic accuracies of $b_{max} \geq 800$ s/mm^2 and $b_{max} < 800$ s/mm^2 in other LF stages. This could be due to that fibrosis is not the only source of altered diffusion properties and ADC values in cirrhotic liver. Previous studies have reported that increased hepatic inflammation degree [26, 29, 30, 32, 43] and liver fat content [57] may lead to reduced ADC as well.

The subgroup analyses also revealed that for cirrhosis patients (F4), IVIM performed statistically significantly better than the conventional DWI model to evaluate liver cirrhosis (Table 5). According to equations defined by Le Bihan et al. [58], *f* is the fraction of microcirculation (perfusion)-related diffusion, *D* is the diffusion parameter of pure molecular diffusion (slow component of diffusion), and *D** that of the perfusion-related diffusion (fast component of diffusion). Liver diffusion combines both pure molecular diffusion and capillary perfusion [51], and with IVIM, pure molecular diffusion can be separated from perfusion-related diffusion with the use of a wide range of b values (including low [<200 s/mm^2] and high [≥200 s/mm^2]). Prior studies [28, 34, 35, 39–41, 45] have shown that IVIM-derived *D** was significantly lower in the fibrotic liver tissues than in the nonfibrotic liver tissues. Moreover, Luciani et al. [12] applied the IVIM model and reported that ADC changes observed in liver cirrhosis were more reflective of a decrease in capillary perfusion than in pure molecular diffusion. However, we failed to identify statistically significant difference between *D** and ADC value in other fibrosis stages; this could be on account of the relatively poor

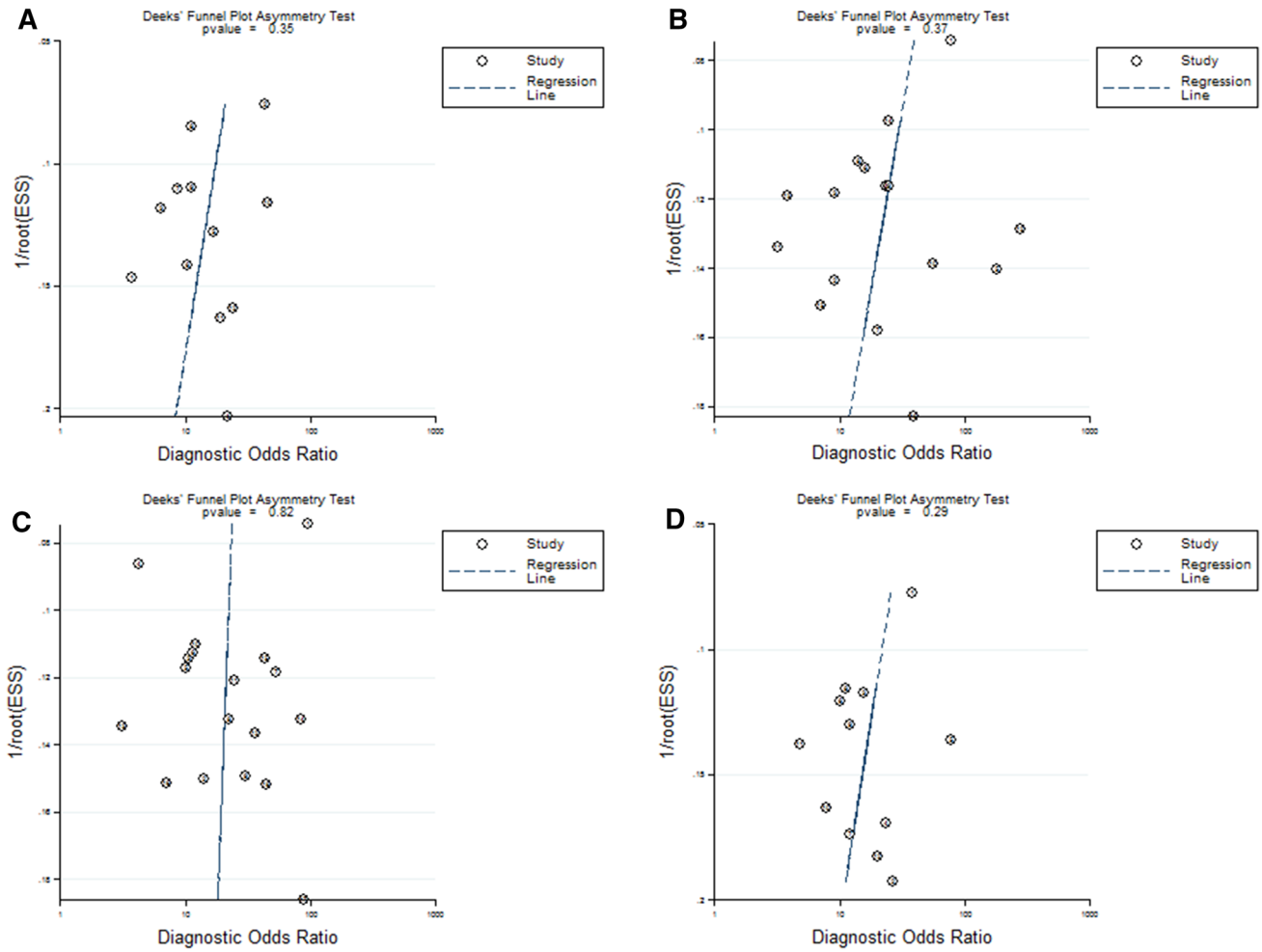

Fig. 4. Funnel plots for DWI in staging LF. The *p* values of the Deeks' funnel plot asymmetry test for LF ≥F1 (**A**), ≥F2 (**B**), ≥F3 (**C**), and ≥F4 (**D**) were 0.35, 0.37, 0.82, and 0.29, respectively, demonstrating no evidence of notable publication bias.

reproducibility of D^* and limited numbers of studies and study population included in our meta-analysis. Therefore, further studies concerning IVIM with better measurement precision and larger study cohort are necessary to further demonstrate the diagnostic performance of IVIM in the staging of LF.

In subgroup analyses, we also compared the effects of two magnetic field strengths: 1.5 and 3.0 T. Our study showed that in liver cirrhosis (F4) group, high-field strength (3.0 T) demonstrated statistically significantly higher sensitivity, specificity, and AUCs compared with low field strength (1.5 T) (Table 5). High-field imaging enables higher signal-to-noise ratio (SNR) [59–61] which either increases spatial resolution or SNR in the ADC maps. However, we failed to identify statistically significant difference between the diagnostic accuracies of 1.5 T and 3.0 T in other fibrosis stages. One possible explanation for this is that in high-strength field, echo-planar imaging results in increased susceptibility artifacts, thus nonecho-planar imaging sequences may optimize the worse image quality and optimize the diagnostic performance at higher field [11]. Therefore, despite the increased availability of 3.0 T imagers, data were still limited on the use of 3.0 T DWI assessing LF, and improved acquisition techniques are required.

However, there are several challenges using DWI to assess LF. First, the acquisition of ADC relies on several imaging parameters including field strength, repetition time, echo time, and b values, thus the reported ADCs of previous studies are variable with considerable overlap between normal and abnormal ranges [13, 24, 30, 46, 47]. Second, images with sufficient quality for reliable quantitative analysis are hard to obtain because DWI is sensitive to susceptibility and motion-related artifacts [13]. Third, a number of potential confounding factors including hepatic perfusion effects, steatosis, edema, hepatic iron, and hepatic necroinflammatory alterations can influence the accurate interpretation of ADC values. Therefore, further studies are required to validate the diagnostic performances of DWI in the assessment of LF and develop standardized DWI methods across different imaging centers.

Our study has limitations. First, the number of studies with high-field strength or IVIM is limited, and the majority of them were overlapped, so it is hard to tell which factor contributed principally for the improved diagnostic performance. Secondly, although QUADAS-2 was adopted to confirm the quality of included studies, there were still many retrospectively designed or unblinded studies included. Therefore, to take full advantage of the benefits of high-field strength and new protocols, prospectively designed large-scale studies specifically addressing those factors are needed in future.

In conclusions, through a larger sample size, comprehensive statistical analysis and the inclusion of high-field scanners and modified IVIM protocol, this study revealed the currently good diagnostic performance of DWI for LF staging, indicated the value of high b value diffusion imaging, and presented the potential role of high-field strength and IVIM for future fibrotic liver imaging.

Acknowledgement. This study was funded by National Natural Science Foundation of China (Grant Number 81471658).

Compliance with ethical standards

Conflict of interest All the authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent For this retrospective type of study, formal consent is not required.

References

1. Tsochatzis EA, Bosch J, Burroughs AK (2014) Liver cirrhosis. Lancet 383:1749–1761
2. Wallace K, Burt AD, Wright MC (2008) Liver fibrosis. Biochem J 411:1–18
3. D'Amico G, Garcia-Tsao G, Pagliaro L (2006) Natural history and prognostic indicators of survival in cirrhosis. A systematic review of 118 studies. J Hepatol 44:217–231
4. Kuramitsu K, Sverdlov DY, Liu SB, et al. (2013) Failure of fibrotic liver regeneration in mice is linked to a severe fibrogenic response driven by hepatic progenitor cell activation. Am J Pathol 183:182–494
5. Marcellin P, Gane E, Buti M, et al. (2013) Regression of cirrhosis during treatment with tenofovir disoproxil fumarate for chronic hepatitis B: a 5-year open-label follow-up study. Lancet 381:468–475
6. Bravo AA, Sheth SG, Chopra S (2001) Liver biopsy. N Engl J Med 344:495–500
7. Bedossa P, Dargere D, Paradis V (2003) Sampling variability of liver fibrosis in chronic hepatitis C. Hepatology 38:1449–1457
8. Castéra L, Foucher J, Bernard PH, et al. (2010) Pitfalls of liver stiffness measurement: a 5-year prospective study of 13,369 examinations. Hepatology 51:828–835
9. Ferraioli G, Tinelli C, Dal Bello B, Zicchetti M, Filice G, Filice C (2012) Accuracy of real-time shear wave elastography for assessing liver fibrosis in chronic hepatitis C: a pilot study. Hepatology 56:2125–2133
10. Huwart L, Sempoux C, Vicaut E, et al. (2008) Magnetic resonance elastography for the noninvasive staging of liver fibrosis. Gastroenterology 135:32–40
11. Taouli B, Koh DM (2010) Diffusion-weighted MR imaging of the liver. Radiology 254:47–66
12. Luciani A, Vignaud A, Cavet M, et al. (2008) Liver cirrhosis: intravoxel incoherent motion MR imaging—pilot study. Radiology 249:891–899
13. Faria SC, Ganesan K, Mwangi I, et al. (2009) MR imaging of liver fibrosis: current state of the art. Radiographics 29:1615–1635
14. Cassinotto C, Feldis M, Vergniol J, et al. (2015) MR relaxometry in chronic liver diseases: comparison of T1 mapping, T2 mapping, and diffusion-weighted imaging for assessing cirrhosis diagnosis and severity. Eur J Radiol 84:1459–1465
15. Wang QB, Zhu H, Liu HL, Zhang B (2012) Performance of magnetic resonance elastography and diffusion-weighted imaging for the staging of hepatic fibrosis: A meta-analysis. Hepatology 56:239–247
16. Poynard T, Bedossa P, Opolon P (1997) Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. Lancet 349:825–832
17. 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
18. Honest H, Khan KS (2002) Reporting of measures of accuracy in systematic reviews of diagnostic literature. BMC Health Serv Res 2:4
19. Arends LR, Hamza TH, van Houwelingen JC, Heijenbrok-Kal MH, HuninkMG Stijnen T (2008) Bivariate random effects meta-analysis of ROC curves. Med Decis Making 28:621–638
20. Higgins JP, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. Stat Med 21:1539–1558
21. Dinnes J, Deeks J, Kirby J, Roderick P (2005) A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. Health Technol Assess 9:1–113
22. Sterne JAC, Harbord RM (2004) Funnel plots in meta-analysis. Stata J 4:127–141
23. Song F, Khan KS, Dinnes J, Sutton AJ (2002) Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. Int J Epidemiol 31:88–95
24. Taouli B, Tolia AJ, Losada M, et al. (2007) Diffusion-weighted MRI for quantification of liver fibrosis: preliminary experience. AJR Am J Roentgenol 189:799–806
25. Feier D, Balassy C, Bastati N, Fragner R, Wrba F, Ba-Ssalamah A (2016) The diagnostic efficacy of quantitative liver MR imaging with diffusion-weighted, SWI, and hepato-specific contrast-enhanced sequences in staging liver fibrosis—a multiparametric approach. Eur Radiol 26:539–546
26. Taouli B, Chouli M, Martin AJ, Qayyum A, Coakley FV, Vilgrain V (2008) Chronic hepatitis: role of diffusion-weighted imaging and diffusion tensor imaging for the diagnosis of liver fibrosis and inflammation. J Magn Reson Imaging 28:89–95
27. Zhou ML, Yan FH, Xu PJ, et al. (2009) Comparative study on clinical and pathological changes of liver fibrosis with diffusion-weighted imaging. Zhonghua Yi Xue Za Zhi 89:1757–1761
28. Parente DB, Paiva FF, Oliveira Neto JA, et al. (2015) Intravoxel incoherent motion diffusion weighted MR imaging at 3.0 T: assessment of steatohepatitis and fibrosis compared with liver biopsy in type 2 diabetic patients. Plos One 10:e0125653
29. Fujimoto K, Tonan T, Azuma S, et al. (2011) Evaluation of the mean and entropy of apparent diffusion coefficient values in chronic hepatitis C: correlation with pathologic fibrosis stage and inflammatory activity grade. Radiology 258:739–748
30. Lewin M, Poujol-Robert A, Boëlle PY, et al. (2007) Diffusion-weighted magnetic resonance imaging for the assessment of fibrosis in chronic hepatitis C. Hepatology 46:658–665
31. Shi Y, Guo QY, Liao W, Ma Y, Qi WX (2010) MR diffusion weighted imaging for quantification of liver fibrosis in patients with chronic viral hepatitis. Chin J Radiol 1:65–69
32. Vaziri-Bozorg SM, Ghasemi-Esfe AR, Khalilzadeh O, et al. (2012) Diffusion-weighted magnetic resonance imaging for diagnosis of liver fibrosis and inflammation in chronic viral hepatitis: the performance of low or high B values and small or large regions of interest. Can Assoc Radiol J 63:304–311
33. Kocakoc E, Bakan AA, Poyrazoglu OK, et al. (2015) Assessment of liver fibrosis with diffusion-weighted magnetic resonance imaging using different b-values in chronic viral hepatitis. Med Princ Pract 24:522–526

34. Wu CH, Ho MC, Jeng YM, et al. (2015) Assessing hepatic fibrosis: comparing the intravoxel incoherent motion in MRI with acoustic radiation force impulse imaging in US. Eur Radiol 25: 3552–3559
35. Ichikawa S, Motosugi U, Morisaka H, Sano K, Ichikawa T, Enomoto N (2015) MRI-based staging of hepatic fibrosis: Comparison of intravoxel incoherent motion diffusion-weighted imaging with magnetic resonance elastography. J Magn Reson Imaging 42:204–210
36. Ding Y, Rao SX, Zhu T, Chen CZ, Li RC, Zeng MS (2015) Liver fibrosis staging using T1 mapping on gadoxetic acid-enhanced MRI compared with DW imaging. Clin Radiol 70:1096–1103
37. Hong Y, Shi Y, Liao W, et al. (2014) Relative ADC measurement for liver fibrosis diagnosis in chronic hepatitis B using spleen/renal cortex as the reference organs at 3 T. Clin Radiol 69:581–588
38. Tokgöz Ö, Unal I, Turgut GG, Yildiz S (2014) The value of liver and spleen ADC measurements in the diagnosis and follow up of hepatic fibrosis in chronic liver disease. Acta Clin Belg 69:426–432
39. Chung SR, Lee SS, Kim N, et al. (2015) Intravoxel incoherent motion MRI for liver fibrosis assessment: a pilot study. Acta Radiol 56:1428–1436
40. Yoon JH, Lee JM, Baek JH, et al. (2014) Evaluation of hepatic fibrosis using intravoxel incoherent motion in diffusion-weighted liver MRI. J Comput Assist Tomogr 38:110–116
41. Chen C, Wang B, Shi D, et al. (2014) Initial study of biexponential model of intravoxel incoherent motion magnetic resonance imaging in evaluation of the liver fibrosis. Chin Med J (Engl) 127:3082–3087
42. Bonekamp D, Bonekamp S, Ou HY, et al. (2014) Assessing liver fibrosis: comparison of arterial enhancement fraction and diffusion-weighted imaging. J Magn Reson Imaging 40:1137–1146
43. Bonekamp S, Torbenson MS, Kamel IR (2011) Diffusion-weighted magnetic resonance imaging for the staging of liver fibrosis. J Clin Gastroenterol 45:885–892
44. Wang Y, Ganger DR, Levitsky J, et al. (2011) Assessment of chronic hepatitis and fibrosis: comparison of MR elastography and diffusion-weighted imaging. AJR Am J Roentgenol 196:553–561
45. Patel J, Sigmund EE, Rusinek H, Oei M, Babb JS, Taouli B (2010) Diagnosis of cirrhosis with intravoxel incoherent motion diffusion MRI and dynamic contrast-enhanced MRI alone and in combination: preliminary experience. J Magn Reson Imaging 31:589–600
46. Sandrasegaran K, Akisik FM, Lin C, et al. (2009) Value of diffusion-weighted MRI for assessing liver fibrosis and cirrhosis. AJR Am J Roentgenol 193:1556–1560
47. Kovač JD, Daković M, Stanisavljević D, et al. (2012) Diffusion-weighted MRI versus transient elastography in quantification of liver fibrosis in patients with chronic cholestatic liver diseases. Eur J Radiol 81:2500–2506
48. Do RK, Chandarana H, Felker E, et al. (2010) Diagnosis of liver fibrosis and cirrhosis with diffusion-weighted imaging: value of normalized apparent diffusion coefficient using the spleen as reference organ. AJR Am J Roentgenol 195:671–676
49. Taouli B, Vilgrain V, Dumont E, Daire JL, Fan B, Menu Y (2003) Evaluation of liver diffusion isotropy and characterization of focal hepatic lesions with two single-shot echoplanar MR imaging sequences: prospective study in 66 patients. Radiology 226:71–78
50. Namimoto T, Yamashita Y, Sumi S, Tang Y, Takahashi M (1997) Focal liver masses: characterization with diffusion-weighted echoplanar MR imaging. Radiology 204:739–744
51. Yamada I, Aung W, Himeno Y, Nakagawa T, Shibuya H (1999) Diffusion coefficients in abdominal organs and hepatic lesions: evaluation with intravoxel incoherent motion echo-planar MR imaging. Radiology 210:617–623
52. Swets JA (1988) Measuring the accuracy of diagnostic systems. Science 240:1285–1293
53. Koinuma M, Ohashi I, Hanafusa K, Shibuya H (2005) Apparent diffusion coefficient measurements with diffusion-weighted magnetic resonance imaging for evaluation of hepatic fibrosis. J Magn Reson Imaging 22:80–85
54. Kim AI, Saab S (2005) Treatment of hepatitis C. Am J Med 118:808–815
55. Ghany MG, Strader DB, Thomas DL, Seeff LB (2009) Diagnosis, management, and treatment of hepatitis C: an update. Hepatology 49:1335–1374
56. Ozkurt H, Keskiner F, Karatag O, Alkim C, Erturk SM, Basak M (2014) Diffusion weighted MRI for hepatic fibrosis: impact of b-value. Iran J Radiol 11:e3555
57. Poyraz AK, Onur MR, Kocakoc E, Ogur E (2012) Diffusion-weighted MRI of fatty liver. J Magn Reson Imaging 35:1108–1111
58. Le Bihan D, Turner R, MacFall JR (1989) Effects of intravoxel incoherent motions (IVIM) in steady-state free precession (SSFP) imaging: application to molecular diffusion imaging. Magn Reson Med 10:324–337
59. Braithwaite AC, Dale BM, Boll DT, Merkle EM (2009) Short- and midterm reproducibility of apparent diffusion coefficient measurements at 3.0-T diffusion-weighted imaging of the abdomen. Radiology 250:459–465
60. van den Bos I, Hussain SM, Krestin GP, Wielopolski PA (2008) Liver imaging at 3.0 T: diffusion-induced black-blood echo-planar imaging with large anatomic volumetric coverage as an alternative for specific absorption rate-intensive echo-train spin-echo sequences: feasibility study. Radiology 248:264–271
61. Lee VS, Hecht EM, Taouli B, Chen Q, Prince K, Oesingmann N (2007) Body and cardiovascular MR imaging at 3.0 T. Radiology 244:692–705