

OPEN

# Accuracy of proton magnetic resonance for diagnosing non-alcoholic steatohepatitis: a meta-analysis

Tae-Hoon Kim<sup>1</sup>, Chang-Won Jeong<sup>1</sup>, Hong Young Jun<sup>1</sup>, ChungSub Lee<sup>1</sup>, SiHyeong Noh<sup>1</sup>, Ji Eon Kim<sup>1</sup>, SeungJin Kim<sup>1</sup> & Kwon-Ha Yoon<sup>1,2\*</sup>

Liver biopsy is the reference standard test to differentiate between non-alcoholic steatohepatitis (NASH) and simple steatosis (SS) in non-alcoholic fatty liver disease (NAFLD), but noninvasive diagnostics are warranted. The diagnostic accuracy in NASH using MR imaging modality have not yet been clearly identified. This study was assessed the accuracy of magnetic resonance imaging (MRI) method for diagnosing NASH. Data were extracted from research articles obtained after a literature search from multiple electronic databases. Random-effects meta-analyses were performed to obtain overall effect size of the area under the receiver operating characteristic(ROC) curve, sensitivity, specificity, likelihood ratios(LR), diagnostic odds ratio(DOR) of MRI method in detecting histopathologically-proven SS(or non-NASH) and NASH. Seven studies were analyzed 485 patients, which included 207 SS and 278 NASH. The pooled sensitivity was 87.4% (95% CI, 76.4–95.3) and specificity was 74.3% (95% CI, 62.4–84.6). Pooled positive LR was 2.59 (95% CI, 1.96–3.42) and negative LR was 0.17 (95% CI, 0.07–0.38). DOR was 21.57 (95% CI, 7.27–63.99). The area under the curve of summary ROC was 0.89. Our meta-analysis shows that the MRI-based diagnostic methods are valuable additions in detecting NASH.

Nonalcoholic fatty liver disease (NAFLD) is clinically ‘silent liver disease’ and most patients with NAFLD are asymptomatic until development of cirrhosis and hepatic decompensation<sup>1</sup>. NAFLD is defined as hepatic fat accumulation ( $\geq 5\%$  of the liver), ballooning hepatocyte degeneration, inflammatory infiltration of polymorphonuclear leukocytes and progressive fibrosis<sup>2</sup>. NAFLD covers a wide spectrum of disease, including simple steatosis, liver inflammation, fibrosis, and cirrhosis<sup>1,3,4</sup>. A subset non-alcoholic steatohepatitis (NASH) is associated with an increased risk for liver cirrhosis and hepatocellular carcinoma (HCC). Overall, 15% to 25% of patients with NASH progress to cirrhotic liver within 10 years, and 4% to 27% of these patients have HCC<sup>5–7</sup>. A “two-hit theory” explains the progression to NASH. The 1st hit is represented by a fat accumulation in the liver, caused by insulin resistance and obesity-induced sensitization<sup>8</sup> and the 2nd hit is represented by the progression of hepatic steatosis, which is induced by various sources of inflammation (including proinflammatory cytokines), oxidative stress, and lipid peroxidation<sup>9</sup>. In clinical practice, liver biopsy is the gold standard method for diagnosing NASH. Only NASH patients at high risk (on the basis of increments of aminotransferase level) undergo a liver biopsy. This may lead to under-diagnosis of NASH<sup>10</sup>. Also, the method has well-known weaknesses with the procedure include the invasive approach, sampling errors, inability to assess the severity of cirrhosis, and complications such as pain, bleeding, infection and rarely death<sup>11,12</sup>. Thus, noninvasive diagnostics are warranted in clinical and imaging modalities could have great potentials to distinguish NASH from NAFLD.

Several methods are used for diagnosing NASH. Considerable effort is underway to identify monitoring strategies that noninvasively diagnose NASH using a variety of techniques such as biochemical tests and imaging techniques. Serum biochemistry which assess the metabolic status, provides information about the liver function<sup>13,14</sup>. However, more than two-thirds of NAFLD patients have normal liver function in the studies, and the parameters for liver function are not significantly correlated with histological findings and are not useful for diagnosing NAFLD<sup>15</sup>. Although several patented serum tests can have sensitivity greater than 80%<sup>16,17</sup>, a recent meta-analysis

<sup>1</sup>Medical Convergence Research Center, Wonkwang University, Iksan, 54538, Republic of Korea. <sup>2</sup>Department of Radiology, Wonkwang University School of Medicine, Iksan, 54538, Republic of Korea. \*email: [khy1646@wku.ac.kr](mailto:khy1646@wku.ac.kr)

study<sup>18</sup> derived from eligible 122 literatures reported that no serum biomarkers based on the pooled sensitivity and specificity revealed good ( $\geq 80\%$ ).

Imaging biomarkers of NASH can noninvasively provide detailed screening information for specific interest or purpose and high diagnostic accuracy suitable for use in large populations. Also, screening large numbers of NASH patients will provide more detailed insight on the mechanisms of pathophysiology and early detection of risk factors in patients. Of the imaging techniques, ultrasound (US) is the first-line imaging test for NAFLD assessment due to its low-cost, wide commercial availability and safety<sup>19,20</sup>. The difference in gray levels on images depends on the acoustic properties (attenuation of acoustic waves, speed of sound, and acoustic impedance) according to different tissue structures. In several studies<sup>21–23</sup>, their sensitivities and specificities for the detection of NAFLD have respectively ranged from 60–94% and 84–95%. And the sensitivities are lower when the degree of NAFLD is mild. Computed tomography (CT) imaging is available because NAFLD decreases the CT attenuation of the liver<sup>24</sup>. A meta-analysis study for detecting the hepatic steatosis reported that the pooled results revealed good sensitivity 81% and specificity 94%. However, these two imaging techniques are not clearly differentiated NASH from simple hepatic steatosis.

Magnetic resonance (MR) technique is a crucial biophysical method for determining the cellular and molecular components<sup>25</sup>. Especially MR imaging (MRI) on the basis of the nuclear MR principles is a robust imaging modality in *in vivo* application to visualize the different image contrasts in normal and abnormal tissues (e.g. simple steatosis, cirrhosis and HCC) for diagnosis<sup>25</sup>. Unlike US and CT, MRI has great merits for screening and staging the NAFLD and/or NASH patients because the method is widely available (as morphology, texture, elastography, strain imaging, diffusion-weighted imaging, perfusion, hepatocellular function and so on) and no ionizing radiation to patients<sup>26,27</sup>. Liver MRI<sup>28</sup> and MR spectroscopy (MRS)<sup>29</sup> has been studied for specific mechanism of human NASH pathophysiology, such as fat accumulation, glucose metabolism, inflammation and oxidative stress. A recent meta-analysis study<sup>30</sup> reported that MRI methods for detecting steatosis or fibrosis in NAFLD had good sensitivity and specificity ( $\geq 80\%$ ). Taking all of these findings into consideration, MRI and MRS could be considered reliable imaging methods for diagnosing NASH instead of liver biopsy<sup>28,29</sup>. However, the diagnostic accuracy of MRI focusing on NASH in large populations has not yet been evaluated sufficiently. Therefore, this study was conducted a systematic review with meta-analyses to identify the diagnostic accuracy of <sup>1</sup>H MRI methods for detecting NASH patients with liver biopsy.

## Result

**Study selection and characteristics of included studies.** Figure 1 shows flowchart showing the process for the inclusion of studies. Following screening, 134 articles were eligible for title and abstract reviews. Inter-observer agreement was very good ( $\kappa = 0.85$ ). After title and abstract screening, we excluded 113 publications as follows: reviews ( $n = 39$ ), interventional study ( $n = 16$ ), irrelevant publications ( $n = 35$ ), not human studies ( $n = 19$ ), case reports/series ( $n = 4$ ). After eligibility screening, 14 articles were excluded with following reasons: no ROC analysis or unclear cut-off ( $n = 10$ ), only fibrosis evaluation ( $n = 1$ ), steatosis severity evaluation ( $n = 1$ ), not clear result ( $n = 1$ ) and non-proton (phosphorus <sup>31</sup>P) MR study ( $n = 1$ ). Finally, 7 studies were included for assessing the diagnostic accuracy of NASH. These studies showed the diagnostic accuracy between biopsy-proven NASH and SS in hepatic fat content ( $n = 2$ ), liver stiffness ( $n = 1$ ), contrast enhanced signal intensity ( $n = 3$ ) and hepatic metabolites ( $n = 1$ ).

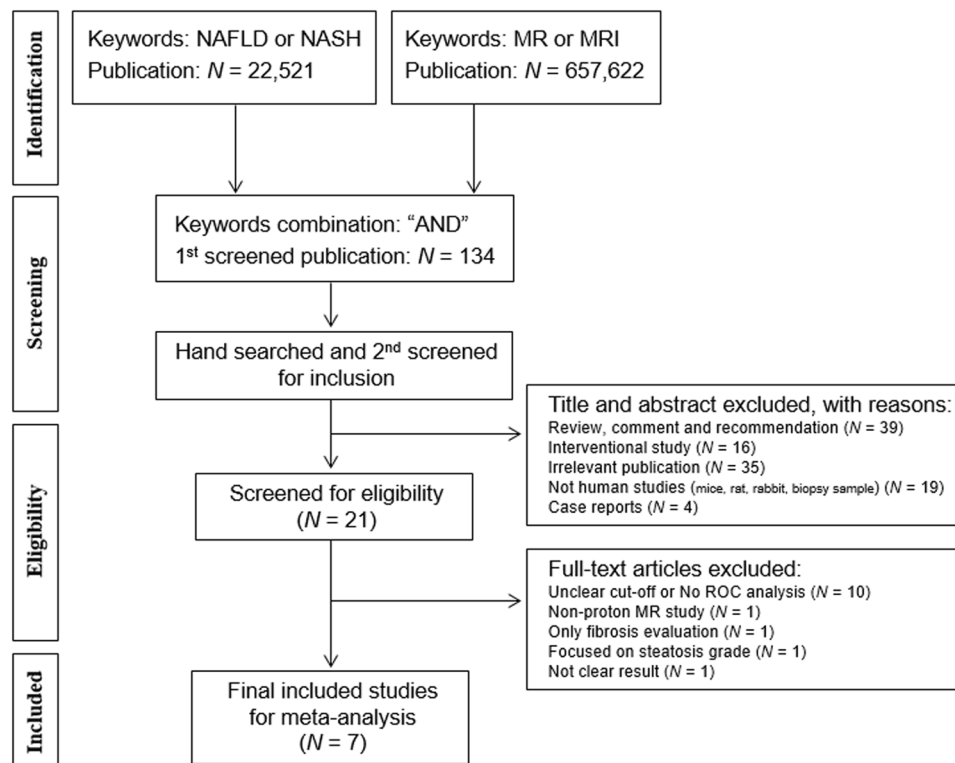
The total number of NAFLD patients in included studies was 485 including of SS 207 and NASH 278. Number of eligible patients analyzed per study ranged from 19 to 190 (median 58). Reported mean age ranged from 38.4 to 55.5 years (mean 48.8 years), while proportion of males ranged from 54% to 79%. Patients were enrolled from the Austria<sup>31</sup>, the United States<sup>32</sup>, Spain<sup>33</sup>, Korea<sup>29</sup>, Netherlands<sup>34</sup>, Japan<sup>35</sup> and the Australia<sup>36</sup>. The four studies (4/7) used 1.5 T MRI scanners (2 GE, 1 Philips and 1 Siemens) and other studies (3/7) used 3.0 T MRI scanner (2 Philips and 1 Siemens). The characteristics of the included studies are detailed in Table 1.

All studies used clinical for the purposes of calculating the test characteristics of MRI to diagnose NASH together with the liver biopsy reference standard. Using the QUADAS-2 assessment tool, all included studies were considered at almost low risk of bias (Fig. 2).

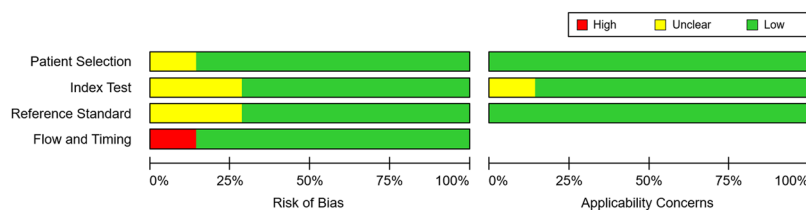
**Diagnostic accuracy for diagnosing NASH.** Figure 3 shows forest plots of the sensitivity and specificity of the included studies. Since inter-study heterogeneity existed ( $I^2 = 79.09\%$  for sensitivity and  $I^2 = 68.89\%$  for specificity), random-effects model was applied to the overall analysis. Sensitivity ranged from 70% to 100% (median 90%) and specificity ranged from 60% to 100% (median 73%). The pooled sensitivity from random-effects regression was 87.4% (95% CI, 76.4–95.3%) and the pooled specificity was 74.3% (95% CI, 62.4–84.6%). Table 2 lists the estimated positive likelihood ratio (LR+), negative LR (LR–) and diagnostic odds ratio (DOR). Pooled LR+ was 2.59 (95% CI, 1.96–3.42) and LR– was 0.17 (95% CI, 0.07–0.38). LR+ and LR– values show weak diagnostic evidence. The DOR was calculated 21.57 (95% CI, 7.27–63.99). Figure 4 shows the summary receiver operating characteristic (SROC) curve for diagnosing NASH in the included studies and its area under the curve (AUC) of SROC curve was 0.8921, indicating very good accuracy.

## Discussion

This study was performed a systematic review and meta-analysis of <sup>1</sup>H MR studies to noninvasively diagnose liver-biopsy proven NASH since 2000. Through the present investigation, 7 studies met eligibility criteria. For the diagnostic accuracy of NASH using MRI, we used the DerSimonian-Laird random-effects model and meta-analytical method for the correlation analysis of sensitivity and specificity. These findings demonstrated that the pooled sensitivity was higher than 80%, but it was significant heterogeneity. Also, publication bias existed for main finding using funnel plot as shown in Fig. 5. The AUC ( $> 0.89$ ) of SROC demonstrated that <sup>1</sup>H MRI is very



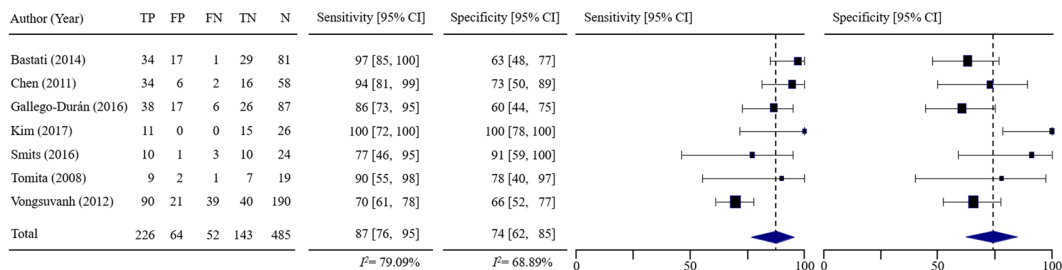
**Figure 1.** Flowchart showing the process for the inclusion of studies. NAFLD: non-alcoholic fatty liver disease; NASH: non-alcoholic steatohepatitis.



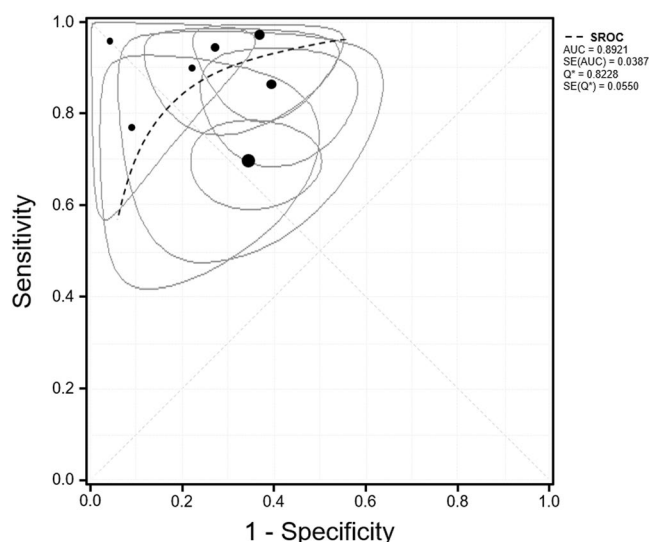
**Figure 2.** QUADAS-2 assessment findings.

First author	Year	Location	No. of patients (n)			Age (Mean ± SD)	Gender (M/F)	Scanner	Reference standard	Scoring system
			Both	SS	NASH					
Bastati	2014	Austria	81	46	35	55.5 ± 13	45/36	3.0T Siemens	Pathology	SAF <sup>d</sup>
Chen	2011	US	58	22	36	51.5 ± ND <sup>a</sup>	ND <sup>a</sup>	1.5 T GE	Pathology	Brunt classification
Gallego-Durán	2016	Spain	87	43	44	50 ± 13	54/33	1.5 T GE & Philips	Pathology	NAS <sup>c</sup>
Kim	2017	Korea	26	15	11	38.4 ± 13	14/12	3.0 T Philips	Pathology	NAS
Smits	2016	Netherlands	24	11	13	54.4 ± 9	17/7	3.0 T Philips	Pathology	NAS
Tomita	2008	Japan	19	9	10	42.0 (38.0–56.5) <sup>b</sup>	15/4	1.5 T GE	Pathology	NAS
Vongsuvan	2012	Australia	190	61	129	49.5 ± 12	110/80	1.5 T Siemens	Pathology	NAS
			485	207	278					

**Table 1.** Basic characteristics of included literatures. F: female; M: male; NASH: nonalcoholic steatohepatitis; SD: standard deviation; and SS: Simple steatosis. <sup>a</sup>ND: not documented; <sup>b</sup>The value presented median (25th percentile–75th percentile); <sup>c</sup>NAS indicated the NAFLD activity score; and <sup>d</sup>SAF system indicated the semiquantitative scoring of steatosis (S), activity (A), and fibrosis (F).



**Figure 3.** Pooled sensitivity and pooled specificity of included studies. The first column includes the last name of the first author for each of the included studies as well as the year of publication in parenthesis, listed in alphabetical order of author’s last name. The next five columns show the number of true positive (TP), false positive (FP), false negative (FN), true negatives (TN) and total number of patients (N) for each of the studies. Sensitivity and specificity are depicted numerically and then graphically as forest plots. In forest plots, each solid square represents an eligible study. The size of the solid square reflects the sample size of each eligible study. Error bars represents 95% CI.



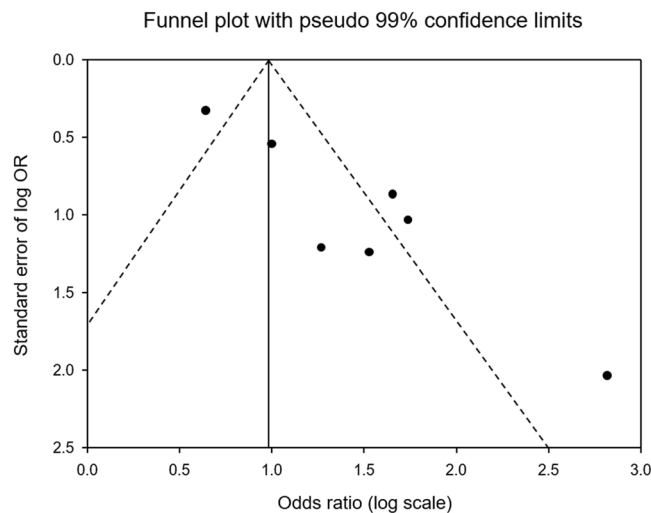
**Figure 4.** Summary receiver operating characteristics (SROC) curve (dashed line). Black circles depict the sensitivity and specificity of individual studies included in this analysis. The ellipse shows the 95% CI for each study. The size of the circle is proportionate to the number of patients enrolled for each study.

	All studies (n = 7)
Sensitivity	0.87 (0.76–0.95)
Specificity	0.74 (0.62–0.85)
LR+	2.59 (1.96–3.42)
LR–	0.17 (0.07–0.38)
DOR	21.57 (7.27–63.99)
PPV	0.81
NPV	0.69

**Table 2.** Pooled test characteristics. The results are reported as point estimates with 95% confidence intervals. LR+ = positive likelihood ratio, LR– = negative likelihood ratio, DOR = diagnostic odds ratio (=LR+/LR–), PPV = positive predictive value, NPV = negative predictive value.

good for the diagnostic accuracy of NASH. Although this study is limited for concrete finding by small number of eligible studies, the <sup>1</sup>H MRI in differential diagnosis of NASH might be considered to use in clinical practice.

In the past decade, there have been efforts for noninvasive diagnosis of NAFLD. Several studies have reported the use of noninvasive imaging techniques including of US, CT, MRI, and MRS as a possible alternative of liver biopsy by evaluating hepatic steatosis<sup>37,38</sup>. These studies have been quantified hepatic fat contents in NAFLD. However, the sensitivities and specificities in these studies vary substantially. Among them, liver MRI has been



**Figure 5.** Funnel plot with pseudo 95% confidence limits for assessment of publication bias.

evaluated for specific mechanism of NASH pathophysiology, such as fat accumulation, glucose metabolism, inflammatory injury and oxidative stress<sup>29,39</sup>. However, an adequately powered, prospective study assessing MRI in terms of reference imaging modality has not yet been performed. Present meta-analysis study can provide the evidence about this gap in knowledge.

Prior to our meta-analysis, several studies have focused on this topic, differential diagnosis of simple steatosis and NASH. Serum biochemical tests for the liver function test are frequently used as indicators of the comprehensive metabolic profile<sup>13,14</sup>. A recent study<sup>18</sup> reported that no serum biomarkers revealed good sensitivity and specificity. Moreover, these parameters for liver function are not definitely correlated with histological findings, as well as are not useful for the diagnosis of NAFLD in clinical practice<sup>15</sup>. In several US studies<sup>21–23</sup>, the sensitivities and specificities have ranged from 60–94% and 84–95%, but it is only for detecting the hepatic steatosis. And their sensitivities are lower when the degree of steatosis is mild. CT imaging studies<sup>21,24</sup> reported that the pooled sensitivity and specificity revealed good (>80%) for detecting the hepatic steatosis. But, all of these studies are not clearly differentiated NASH from simple hepatic steatosis. Present study was performed a meta-analysis from seven studies comprising totally 485 biopsy-proven adult patients. Pooled LR+ (2–5) and LR– (<0.2) indicated that there is weak diagnostic evidence in the differentiation of both NASH and SS patients. DOR showed higher diagnostic accuracy of the MRI test as a NASH patient relative to that in a simple steatosis without inflammation.

With regard to data quality, our eligible studies were judged at low risk. This meta-analysis included an explanation for data extraction, forest plots for pooling test, and the evaluation for the uncertainty around point estimates. The QUADAS-2 domains (consisting of subject selection, index test, reference standard test, and flow & timing) showed low publication bias. Thus, our meta-analysis would be provided useful information in differentiating NASH from simple hepatic steatosis, unlike previous reports of the diagnostic accuracy of laboratory test<sup>18</sup>, US<sup>21–23</sup> and CT<sup>24</sup>. Heterogeneity among eligible studies may point out the potential bias or variation, therefore it is essential for any meta-analyses that sources of heterogeneity are tested to identify the potential risk factors by interstudy differences in selection of patient (age, NASH severity, gender and so on), the cutoff-point values for differential diagnosis, clinical conditions (type of drug, initial management, drug dosage, treatment duration and etc.) and imaging setting (type of scanners/pulse sequences) or any combinations with the sources. Also, heterogeneity can be caused by defects in the design of study<sup>40</sup>. Moreover, heterogeneity on diagnostic tests might be influenced by the reliability in measurement and/or duration of follow-up<sup>41</sup>. Drawback in report, analysis and interpretation from results occurred frequently in the assessment of the included literatures. To overcome the problem of heterogeneity, we suggest the standardized study protocol/design, prospective, large-scale cohort studies to clarify unmet needs in clinical interests.

Our findings demonstrated that <sup>1</sup>H MRI could be a rational alternative to liver biopsy for diagnosing the NASH in hospitals. Especially, <sup>1</sup>H MRI examinations are noticeably useful for screening the early identification of patients at high risk of NASH and reducing the liver biopsy. However, this study included a number of limitations. First, a variety of aspects, as the different populations used (broad ranges of age and diffuse lesion), the diverse number of contents evaluated (hepatic fat, stiffness, contrast-enhanced signal intensity, and metabolites), and the use of diverse cut-off values for the same content, may give rise to a broad range of results in diagnostic accuracy of NASH. Moreover, another concern refers to the heterogeneity of studies, given the potential differences in diverse methods for collection, transportation and storage of biopsy specimen, pre- & post-processing of MR data, and in the uncertainties in the data stability. Information about these issues is still limited. These factors may lead to disturb precise pooling results in the meta-analysis and cause variations in the results for the imaging marker of NASH. Second, the included case-control studies showed a relatively high prevalence of NASH presenting to the NAFLD with severe symptom at high risk when compared to the general population, concerning NASH prevalence. To enroll the patients with a high probability of having NASH through pre-testing, this is likely to use the scheduled liver biopsy in some studies. This may result in overestimation of diagnostic accuracy

(spectrum bias) and restrict the external validity of SROC curve generalized to the target population. Also, patient registry using convenient sampling methods was not sufficiently explained to clarify probability in false-negative results. This may give rise to spectrum bias and information bias in meta-analysis. To overcome this issue, further study should enroll a large-scale relevant cohort to reflect the general population. Third, the difference of MRI protocols in the eligible studies was not sufficiently considered for this study. Therefore, the use of a reference standard imaging would be useful for directly comparing the performance of test and clearly elucidating the efficacy of MRI. Fourth, although our study only included pathologically-proven NASH/NAFLD patients, it is required to focus more on hepatic inflammation and hepatocellular injury (ballooning). Further study is needed to clarify MRI-based methods for quantifying these more important histologic features (inflammation and ballooning).

In conclusion, this meta-analysis summarizes the evidence about the accuracy of <sup>1</sup>H MRI examinations for noninvasively diagnosing NASH, with an overall AUC of 0.89. The pooled results reveal good sensitivity of 87.4% and significant diagnostic performance (LR+ 2.59, LR- 0.17 and DOR 21.57). MRI-based diagnostic methods are valuable additions for detecting NASH. The <sup>1</sup>H MRI examination would be useful for screening of early identification of NASH and reducing of liver biopsy in clinical. However, it is limited for exclusively use in clinical applications. Replication studies and more standardized study designs are required to be an efficient imaging biomarker for NASH diagnosis in clinical practice.

## Materials and Methods

**Literature search.** This study was conducted meta-analyses and systematic reviews according to the Cochrane Collaboration guidelines. A systematic search was conducted in Cochrane library, EMBASE and Medline database (January 2000 until December 2017), with the following keywords based on Medical Subject Headings terms: fatty liver, nonalcoholic fatty liver disease (NAFLD), NAFL or nonalcoholic steatohepatitis (NASH), in any combination with MRI examinations. Each search was restricted to English language publications (Fig. 1).

**Selecting study and extracting data.** Following deduplication, two authors (T.-H.K., C.-W.J.) independently screened the titles and abstracts of studies. Full-text articles were then reviewed for potential eligibility by the authors. Disagreement between both authors at any viewpoint for analysis was mediated by a 3rd reviewer (K.-H.Y.). These three reviewers extracted in duplicate data from eligible studies using a pre-designed form. Finally, we extracted data to reconstruct 2 × 2 tables for each MRI examination for every individual study.

**Inclusion and exclusion criteria for eligibility.** To be eligible for inclusion, MRI studies were required to diagnose NASH with liver biopsy as the reference standard. Studies included the following values for clinical diagnosis of NASH: sensitivity, specificity, diagnostic odds ratio (DOR), area under the receiver-operating characteristic (AUROC) or diagnostic accuracy. We excluded the following studies: animal, case reports or series report, *in vitro* biopsy sample, non-proton nuclei, reviews and irrelevant publication (alcoholic liver disease; other chronic liver diseases or liver diseases with two or more mixed etiologies except NAFLD). Also, studies focusing only on steatosis and/or fibrosis in NAFLD or not differentiating NASH from simple steatosis were excluded.

**Assessment of methodological quality.** The quality of included studies was evaluated based on the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) criteria<sup>42</sup>. The QUADAS-2 tool consists of four key domains: patient selection; index test; reference standard; and flow and timing. The domains are evaluated in terms of risk of bias and applicability. Each included article independently assessed by two authors using this tool and disagreement between both authors was resolved by a consensus.

**Statistical analysis and data synthesis.** In order to evaluate the diagnostic accuracy of MRI in biopsy-proven NASH, this study was used a bivariate random effects model for the meta-analysis of sensitivity and specificity. All analyses were carried out using Meta-DiSc<sup>®</sup> software (ver. 1.4; Clinical Biostatistics Unit, Ramón y Cajal Hospital, Madrid, Spain) and meta-analysis of diagnostic accuracy (MADA) package on the R program language (R Foundation, Vienna, Austria). To consider the overall effect size for the pooled analyses, we calculated the inverse variance-weighted average adjusted for effect size of each study.

The procedure of meta-analysis accounts for interstudy variability and possible correlation between sensitivity and specificity. Both sensitivity and specificity were estimated to 2 × 2 tables (true positive, false positive, false negative and true negative results) with 95% confidence intervals (CI). We identified the average operating point and computed average sensitivity, specificity, positive LR (LR+), negative LR (LR-) and DOR (DOR = LR+/LR-). To calculate the pooled sensitivity and specificity, we added 0.5 to all zero cells in two-by-two tables when the value of false positive or false negative is zero in each study. LR+ is the best indicator for ruling-in diagnosis. The higher the LR+ the test is more indicative of a disease. LR- is a good indicator for ruling-out the diagnosis<sup>43</sup>. The evidence of diagnostic tests on the basis of LR+ and LR- values is interpreted as conclusive diagnostic evidence (LR+ >10, LR- <0.01), strong diagnostic evidence (LR+ >5, LR-, <0.2), weak diagnostic evidence (LR+ 2–5, LR- 0.2–0.5) and negligible evidence (LR+ 1–2, LR- 0.5–1)<sup>43</sup>. Here, the DOR summarized the diagnostic accuracy of an index test as a single value, describing how many times higher the odds are of obtaining a test positive result in a NASH rather than a simple steatosis<sup>44</sup>.

The random effects model meta-analyses were chosen for two reasons. The 1<sup>st</sup> reason was the necessity for a pooling effect of the sizes derived from different diagnostic contents & imaging methods and clinical conditions of NASH, and differentiation cut-off value of MRI used in each study. The 2<sup>nd</sup> reason, the selection of random effects model was due to the statistical heterogeneity of the outcome data. We fitted the DerSimonian and Laird random effects model to depict SROC curves. To assess inter-study inconsistency, we tested by I<sup>2</sup> index based on Chi<sup>2</sup> tests, with I<sup>2</sup> values higher than 50% or *p* < 0.10 respectively representing high heterogeneity<sup>45</sup>. An area under the curve (AUC) of SROC indicates as follows: 0.9–1.0 = excellent; 0.8–0.9 = very good; 0.7–0.8 = good;

0.6–0.7 = sufficient; and 0.5–0.6 = bad<sup>44</sup>. To assess publication bias, inverted funnel plots of the logarithmical OR from each study were plotted against the sample size<sup>46</sup>. All the data are expressed as weighted effect sizes with 95% CI (Supplementary Figures about effect size can be found in supporting information file).

Received: 8 October 2018; Accepted: 28 September 2019;

Published online: 18 October 2019

## References

- Milic, S. & Stimac, D. Nonalcoholic fatty liver disease/steatohepatitis: epidemiology, pathogenesis, clinical presentation and treatment. *Dig. Dis.* **30**, 158–162 (2012).
- Yatsuji, S. *et al.* Clinical features and outcomes of cirrhosis due to non-alcoholic steatohepatitis compared with cirrhosis caused by chronic hepatitis C. *J Gastroenterol. Hepatol.* **24**, 248–254 (2009).
- Lonardo, A., Byrne, C. D., Caldwell, S. H., Cortez-Pinto, H. & Targher, G. Global epidemiology of nonalcoholic fatty liver disease: Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* **64**, 1388–1389 (2016).
- Farrell, G. C., Wong, V. W. & Chitturi, S. NAFLD in Asia—as common and important as in the West. *Nat. Rev. Gastroenterol. Hepatol.* **10**, 307–318 (2013).
- McCullough, A. J. Pathophysiology of nonalcoholic steatohepatitis. *J Clin. Gastroenterol.* **40**(Suppl 1), S17–29 (2006).
- Wong, V. W. *et al.* Disease progression of non-alcoholic fatty liver disease: a prospective study with paired liver biopsies at 3 years. *Gut* **59**, 969–974 (2010).
- Ascha, M. S. *et al.* The incidence and risk factors of hepatocellular carcinoma in patients with nonalcoholic steatohepatitis. *Hepatology* **51**, 1972–1978 (2010).
- Bezy, O. *et al.* PKCdelta regulates hepatic insulin sensitivity and hepatosteatosis in mice and humans. *J Clin. Invest.* **121**, 2504–2517 (2011).
- Csak, T. *et al.* Fatty acid and endotoxin activate inflammasomes in mouse hepatocytes that release danger signals to stimulate immune cells. *Hepatology* **54**, 133–144 (2011).
- Verdam, F. J. *et al.* Non-alcoholic steatohepatitis: a non-invasive diagnosis by analysis of exhaled breath. *J Hepatol.* **58**, 543–548 (2013).
- Howlett, D. C., Drinkwater, K. J., Lawrence, D., Barter, S. & Nicholson, T. Findings of the UK national audit evaluating image-guided or image-assisted liver biopsy. Part II. Minor and major complications and procedure-related mortality. *Radiology* **266**, 226–235 (2013).
- Ratziu, V. *et al.* Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology* **128**, 1898–1906 (2005).
- Chang, Y., Ryu, S., Sung, E. & Jang, Y. Higher concentrations of alanine aminotransferase within the reference interval predict nonalcoholic fatty liver disease. *Clin. Chem.* **53**, 686–692 (2007).
- Zhang, J. *et al.* Association between serum free fatty acid levels and nonalcoholic fatty liver disease: a cross-sectional study. *Sci. Rep.* **4**, 5832 (2014).
- Dowman, J. K., Tomlinson, J. W. & Newsome, P. N. Systematic review: the diagnosis and staging of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis. *Aliment. Pharmacol. Ther.* **33**, 525–540 (2011).
- Alkhoury, N. & Feldstein, A. E. Noninvasive diagnosis of nonalcoholic fatty liver disease: Are we there yet? *Metabolism* **65**, 1087–1095 (2016).
- Hadizadeh, F., Faghihimi, E. & Adibi, P. Nonalcoholic fatty liver disease: Diagnostic biomarkers. *World J Gastrointest. Pathophysiol.* **8**, 11–26 (2017).
- Verhaegh, P. *et al.* Noninvasive Tests Do Not Accurately Differentiate Nonalcoholic Steatohepatitis From Simple Steatosis: A Systematic Review and Meta-analysis. *Clin. Gastroenterol. Hepatol.* **16**, 837–861 (2018).
- Liao, Y. Y. *et al.* Multifeature analysis of an ultrasound quantitative diagnostic index for classifying nonalcoholic fatty liver disease. *Sci. Rep.* **6**, 35083 (2016).
- Khov, N., Sharma, A. & Riley, T. R. Bedside ultrasound in the diagnosis of nonalcoholic fatty liver disease. *World J Gastroenterol.* **20**, 6821–6825 (2014).
- Schwenzer, N. F. *et al.* Non-invasive assessment and quantification of liver steatosis by ultrasound, computed tomography and magnetic resonance. *J Hepatol.* **51**, 433–445 (2009).
- Chen, C. H. *et al.* The accuracy of sonography in predicting steatosis and fibrosis in chronic hepatitis C. *Dig. Dis. Sci.* **53**, 1699–1706 (2008).
- Gomercic, M., Duvnjak, M. & Barsic, N. Ultrasonography in the diagnosis of nonalcoholic fatty liver disease. *Acta Med. Croatica* **63**(Suppl 3), 1–3 (2009).
- Iwasaki, M. *et al.* Noninvasive evaluation of graft steatosis in living donor liver transplantation. *Transplantation* **78**, 1501–1505 (2004).
- Mandal, P. K. *In vivo* proton magnetic resonance spectroscopic signal processing for the absolute quantitation of brain metabolites. *Eur J Radiol* **81**, e653–664 (2012).
- Petitclerc, L., Sebastiani, G., Gilbert, G., Cloutier, G. & Tang, A. Liver fibrosis: Review of current imaging and MRI quantification techniques. *J Magn. Reson. Imaging* **45**, 1276–1295 (2017).
- Dyvorne, H. A. *et al.* Prospective comparison of magnetic resonance imaging to transient elastography and serum markers for liver fibrosis detection. *Liver Int.* **36**, 659–666 (2016).
- Kim, T. H. *et al.* Noninvasive Differential Diagnosis of Liver Iron Contents in Nonalcoholic Steatohepatitis and Simple Steatosis Using Multiecho Dixon Magnetic Resonance Imaging. *Acad. Radiol.* **26**, 766–774 (2019).
- Kim, T. H. *et al.* Hepatic Alanine Differentiates Nonalcoholic Steatohepatitis From Simple Steatosis in Humans and Mice: A Proton MR Spectroscopy Study With Long Echo Time. *J Magn. Reson. Imaging* **46**, 1298–1310 (2017).
- Wang, X. M., Zhang, X. J. & Ma, L. Diagnostic performance of magnetic resonance technology in detecting steatosis or fibrosis in patients with nonalcoholic fatty liver disease: A meta-analysis. *Medicine (Baltimore)* **97**, e10605 (2018).
- Bastati, N. *et al.* Noninvasive differentiation of simple steatosis and steatohepatitis by using gadoxetic acid-enhanced MR imaging in patients with nonalcoholic fatty liver disease: a proof-of-concept study. *Radiology* **271**, 739–747 (2014).
- Chen, J. *et al.* Early detection of nonalcoholic steatohepatitis in patients with nonalcoholic fatty liver disease by using MR elastography. *Radiology* **259**, 749–756 (2011).
- Gallego-Duran, R. *et al.* Imaging biomarkers for steatohepatitis and fibrosis detection in non-alcoholic fatty liver disease. *Sci. Rep.* **6**, 31421 (2016).
- Smits, L. P. *et al.* Noninvasive Differentiation between Hepatic Steatosis and Steatohepatitis with MR Imaging Enhanced with USPIOs in Patients with Nonalcoholic Fatty Liver Disease: A Proof-of-Concept Study. *Radiology* **278**, 782–791 (2016).
- Tomita, K. *et al.* Evaluating the severity of nonalcoholic steatohepatitis with superparamagnetic iron oxide-enhanced magnetic resonance imaging. *J Magn. Reson. Imaging* **28**, 1444–1450 (2008).
- Vongsuvanhan, R., George, J., McLeod, D. & van der Poorten, D. Visceral adiposity index is not a predictor of liver histology in patients with non-alcoholic fatty liver disease. *J Hepatol.* **57**, 392–398 (2012).

37. Bohte, A. E., van Werven, J. R., Bipat, S. & Stoker, J. 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 (2011).
38. Lee, S. S. & Park, S. H. Radiologic evaluation of nonalcoholic fatty liver disease. *World J Gastroenterol.* **20**, 7392–7402 (2014).
39. Reeder, S. B., Cruite, I., Hamilton, G. & Sirlin, C. B. Quantitative assessment of liver fat with magnetic resonance imaging and spectroscopy. *J Magn. Reson. Imaging* **34**, 729–749 (2011).
40. Lijmer, J. G., Bossuyt, P. M. & Heisterkamp, S. H. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat. Med.* **21**, 1525–1537 (2002).
41. Glasziou, P. P. & Sanders, S. L. Investigating causes of heterogeneity in systematic reviews. *Stat. Med.* **21**, 1503–1511 (2002).
42. Whiting, P. F. *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann. Intern. Med.* **155**, 529–536 (2011).
43. Deeks, J. J. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. *BMJ* **323**, 157–162 (2001).
44. Glas, A. S., Lijmer, J. G., Prins, M. H., Bonsel, G. J. & Bossuyt, P. M. The diagnostic odds ratio: a single indicator of test performance. *J Clin. Epidemiol.* **56**, 1129–1135 (2003).
45. Bell, L. N. *et al.* Serum proteomics and biomarker discovery across the spectrum of nonalcoholic fatty liver disease. *Hepatology* **51**, 111–120 (2010).
46. Targher, G., Byrne, C. D., Lonardo, A., Zoppini, G. & Barbui, C. Non-alcoholic fatty liver disease and risk of incident cardiovascular disease: A meta-analysis. *J Hepatol.* **65**, 589–600 (2016).

## Acknowledgements

This study was supported by the grants of the National Research Foundation of Korea (NRF) (2016M3A9A7918501) and the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare (HI18C1216).

## Author contributions

All authors contributed to this work. Study design: T.-H.K., C.-W.J., K.-H.Y. Screening of abstracts and articles: T.-H.K., C.-W.J. Data extraction: T.-H.K., H.Y.J., C.S.L., S.H.N., J.-E.K., S.J.K. Compile supplementary tables: T.-H.K., C.-W.J., H.Y.J., C.S.L., S.H.N., J.-E.K., S.J.K., K.-H.Y. Constitute figures: T.-H.K., C.-W.J., H.Y.J., C.S.L., S.H.N., J.-E.K., S.J.K., K.-H.Y. Data analysis: T.-H.K., C.-W.J., H.Y.J., C.S.L., S.H.N., J.-E.K., S.J.K., K.-H.Y. Writing: T.-H.K. Critically reviewing manuscript: T.-H.K., C.-W.J., K.-H.Y.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-019-51302-w>.

**Correspondence** and requests for materials should be addressed to K.-H.Y.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019