The diagnostic accuracy of US, CT, MRI and (1)H-MRS for the evaluation of hepatic steatosis compared with liver biopsy: a meta-analysis
Bohte, A.E.; van Werven, J.R.; Bipat, S.; Stoker, J.

Published in:
European Radiology

DOI:
10.1007/s00330-010-1905-5

Citation for published version (APA):
The diagnostic accuracy of US, CT, MRI and \(^1\)H-MRS for the evaluation of hepatic steatosis compared with liver biopsy: a meta-analysis

ANNELOES E. BOHTE · JOCHEN R. VAN WERVEN · SHANDRA BIPAT · JAAP STOKER

Received: 9 April 2010 / Revised: 24 May 2010 / Accepted: 14 June 2010 / Published online: 31 July 2010 © The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract
Objective To meta-analyse the diagnostic accuracy of US, CT, MRI and \(^1\)H-MRS for the evaluation of hepatic steatosis.

Methods From a comprehensive literature search in MEDLINE, EMBASE, CINAHL and Cochrane (up to November 2009), articles were selected that investigated the diagnostic performance imaging techniques for evaluating hepatic steatosis with histopathology as the reference standard. Cut-off values for the presence of steatosis on liver biopsy were subdivided into four groups: (1) >0, >2 and >5% steatosis; (2) >10, >15 and >20%; (3) >25, >30 and >33%; (4) >50, >60 and >66%. Per group, summary estimates for sensitivity and specificity were calculated. The natural-logarithm of the diagnostic odds ratio (lnDOR) was used as a single indicator of test performance.

Results 46 articles were included. Mean sensitivity estimates for subgroups were 73.3–90.5% (US), 46.1–72.0% (CT), 82.0–97.4% (MRI) and 72.7–88.5% (\(^1\)H-MRS). Mean specificity ranges were 69.6–85.2% (US), 88.1–94.6% (CT), 76.1–95.3% (MRI) and 92.0–95.7% (\(^1\)H-MRS). Overall performance (lnDOR) of MRI and \(^1\)H-MRS was better than that for US and CT for all subgroups, with significant differences in groups 1 and 2.

Conclusion MRI and \(^1\)H-MRS can be considered techniques of choice for accurate evaluation of hepatic steatosis.

Keywords Hepatic steatosis · Diagnostic accuracy · Magnetic resonance imaging · Magnetic resonance spectroscopy · Computed tomography · Ultrasonography

Introduction

The prevalence of hepatic steatosis is increasing rapidly worldwide. This is largely attributed to the association with obesity and insulin resistance in non-alcoholic fatty liver disease (NAFLD) [1, 2].

Detection and quantification of hepatic steatosis is clinically important. In NAFLD, steatosis is the hepatic manifestation of the metabolic syndrome and the earliest biomarker for the development of liver fibrosis in the more severe condition of non-alcoholic steatohepatitis (NASH). Early diagnosis and treatment of NASH can prevent the potential development of cirrhosis and hepatocellular carcinoma (HCC) [3–5]. In hepatitis C, steatosis is associated with more severe fibrosis and rapid disease progression [6, 7]. In liver transplantation surgery, the presence of steatosis impairs the regenerative capacity of the liver in both donor and recipient [8, 9].

Liver biopsy remains the reference test for the evaluation of hepatic steatosis, despite well-established drawbacks regarding its invasiveness and sampling error due to small sample size and inter-observer variability [10].

Many studies have focused on the role of imaging techniques as a non-invasive alternative to liver biopsy for detecting and quantifying hepatic steatosis [11–13]. The reported sensitivities and specificities between different imaging techniques and between different studies investi-
gating the same technique vary substantially. Although magnetic resonance spectroscopy ($^1$H-MRS)—generally considered the best technique—is increasingly used as a reference standard instead of liver biopsy, no evidence-based consensus currently exists on this topic.

The purpose of this systematic review therefore was to summarise the available literature on the accuracy of ultrasound (US), computed tomography (CT), magnetic resonance imaging (MRI) and $^1$H-MRS for the evaluation of hepatic steatosis with histopathology as the reference test. Subsequently, we aimed to identify the most accurate technique by meta-analysis.

Materials and methods

Literature search and study selection

We searched the MEDLINE (January 1966–November 2009), EMBASE (January 1980–November 2009), CINAHL and Cochrane databases without language restrictions with the assistance of an experienced clinical librarian. We combined Medical Subject Headings (MeSH) terms and accompanying entry terms for the patient group (patients with hepatic steatosis) and the index test (US, CT, MRI, $^1$H-MRS). The search strategy is described in detail in Online Resource 1.

Two reviewers (A.B. and J.v.W) read the titles and abstracts of all the articles obtained to select potentially relevant papers (original papers that addressed the diagnostic accuracy of US, CT, MRI or $^1$H-MRS for detecting hepatic steatosis in humans with histopathology as the reference test in ≥10 individuals). The reference lists of selected papers and of narrative reviews were screened for search completion.

The full texts of potentially relevant papers were reviewed for inclusion by the same reviewers independently. Inclusion criteria were: (a) hepatic steatosis was evaluated with US, CT, MRI and/or $^1$H-MRS; (b) imaging techniques met the minimum technical requirements of grey scale and real time for US; ≤120 kV for CT and ≥1 Tesla for MRI; (c) histopathology as the reference test; (d) evaluation of ≥10 human individuals; (e) criteria for a positive index test were clearly explained; (f) examination method of steatosis on liver biopsy was clearly explained; (g) data on diagnostic accuracy were reported. Exclusion criteria were: (a) duplicate publication; (b) reporting of combined data for different imaging techniques or data on the single technique could not be extracted; (c) no original research. Papers were not blinded with regard to authors’ names, affiliations or journal. The reviewers resolved all disagreements about inclusion and data extraction by consensus after face-to-face discussion.

Data extraction

From the articles included, the reviewers (A.B. and J.v.W.) independently recorded data using a standardised form. Papers were translated if necessary.

Methodological quality Methodological quality was assessed based on the Quality Assessment of Studies of Diagnostic Accuracy included in Systematic Reviews (QUADAS) guidelines [14]. To be reasonably sure that the condition of the liver did not change between the two tests we chose a period of 1 month as a quality indicator [15, 16]. Additionally, we noted whether the study design was prospective or retrospective.

Patient characteristics For each study, we extracted data on (a) sample size; (b) male-female ratio; (c) patient age; (d) body mass index (BMI) and (e) patient spectrum.

Imaging features and evaluation For each imaging technique we recorded (a) the number of patients; (b) the criteria used for steatosis evaluation and (c) the cut-off values used. We noted whether cut-off values were defined prospectively or retrospectively. Additionally, for US we noted the type and frequency of the probe(s) used. For CT we noted: (a) the type of CT and (b) the imaging parameters (kV and mAs). For MRI, we noted: (a) magnetic field strength; (b) imaging sequence; (c) imaging parameters used; (d) whether breath holds were used; and (e) correction for T2* effects. For $^1$H-MRS, we noted: (a) magnetic field strength; (b) imaging sequence; (c) imaging parameters; (d) voxel size; (e) whether breath holds were used and (f) correction for T1 or T2 effects.

Reference test For liver biopsy we included data on: (a) the number of patients biopsied; (b) cut-off values for steatosis grades and whether the presence of (c) fibrosis (d) inflammation and (e) iron was evaluated in the biopsy specimen.

Data for calculation of diagnostic accuracy We extracted available data on true-positives (TP), false-negatives (FN), false-positives (FP) and true-negatives (TN) for detecting steatosis with the selected imaging technique to construct $2 \times 2$ contingency tables. Available $3 \times 3$ and $4 \times 4$ tables were dichotomised. Many different cut-off values for positive results (steatosis present) on liver biopsy were compared to the imaging techniques. We therefore grouped accuracy results from cut-off values that were almost equal into four subgroups to enable meta-analysis:

Group 1: Cut-off values of >0%, >2% and >5% steatosis on biopsy;
Group 2: Cut-off values of >10%, >15% and >20% steatosis on biopsy;
Group 3: Cut-off values of >25%, >30% and >33% steatosis on biopsy and the qualitative designation of “moderate or severe” steatosis;
Group 4: Cut-off values of >50%, >60% and >66% steatosis on biopsy and the qualitative designation of “severe” steatosis.

During analysis we corrected for dependent data such as results presented for different readers or for multiple imaging techniques within one study population. For CT and MRI, we did not use data obtained by subjective visual evaluation of examinations for analysis. If raw data in terms of 2×2 tables were unavailable, we attempted to contact authors for completion or verification of data.

Data analysis

We performed bivariate random-effects analysis for the pooled sensitivities and specificities per cut-off value group for each imaging technique [17]. In this analysis, the logit-transformed sensitivities and logit-transformed specificities from individual studies in a meta-analysis are assumed to follow a bivariate normal distribution around a mean logit sensitivity and a mean logit specificity. After antilogit transformation, summary estimates of sensitivity and specificity with 95% confidence intervals were calculated. Additionally, we calculated the natural logarithm (ln) of the diagnostic odds ratio (DOR): [logit sensitivity + logit specificity]. The DOR is a single indicator of test performance [18]. A higher lnDOR value indicates a better discriminatory test performance. If the lnDOR is not significantly different from 0, a test does not discriminate between patients with the disorder and those without it. We performed z-tests to compare sensitivities, specificities and lnDORS between imaging techniques. A p value<0.05 was considered statistically significant. All meta-analyses were performed with SAS statistical software (version 9.1, SAS institute, Cary, NC, USA).

Results

Literature search and study selection

The literature search yielded 6992 unique references (Fig. 1). The reviewers selected 179 potentially relevant articles after reading the titles and abstracts of which 46 papers were finally included [15, 19–63]. Twenty-eight evaluated US, twelve evaluated CT, ten evaluated MRI and five evaluated 1H-MRS. Eight papers compared two imaging techniques [22, 23, 43, 47, 54, 58, 61, 62] and one evaluated three techniques within the same population [56] (Table 1). Two studies were published in German [28, 45] and the remaining in English.

Data extraction

Methodological quality An overview of the results is given in Fig. 2. In general, 13 out of 46 (28%) studies fulfilled at least 10 of the 13 methodological criteria [19, 20, 25, 29, 33, 34, 36, 38, 41, 44, 48, 49, 52]. A complete table of individual study scores is available upon request from the authors.

Patient characteristics The 46 papers comprised 4715 patients with a median study size of 81 patients (range 20–589). Mean age reported by 37 studies was 44.5 years (range 11–89). Mean BMI reported by 18 studies was 26.6 kg/m² (range 15–54 kg/m²). The male-to-female ratio reported in 40 studies was 1.62:1. Potential living liver donors constituted 34% of the total population (1593/4715). If specified, the disease spectrum comprised most frequently chronic hepatitis C (n=1040) and NAFLD/NASH (n=710). See also Table 1.

Imaging features and evaluation US imaging features are outlined in Online Resource 2. More than half (15/28) of the studies used the widely accepted criteria for subjective visual steatosis evaluation of bright liver with increased liver-kidney contrast; blurring of intrahepatic vessels and diaphragm and loss of echoes of posterior hepatic segments [57, 60]. Two studies evaluated quantitative methods to assess liver steatosis [30, 59].

CT imaging features are shown in Online Resource 3. Two papers evaluated both contrast-enhanced CT and unenhanced CT [22, 38]. All other papers evaluated unenhanced CT. Average Hounsfield Units (HU) in selected regions of interests (ROIs) from the liver were compared with average HU values of ROIs from the spleen. The spleen was used as an internal reference in all the papers included, either by measuring the liver-minus-spleen attenuation value (L-S) or the liver-to-spleen ratio (L/S). Two studies also evaluated steatosis by subtracting hepatic blood attenuation from the total hepatic attenuation using an algorithm. Four papers had defined cut-off values prospectively [22, 43, 58, 61]. Six papers defined optimal cut-off values retrospectively [15, 35, 38, 44, 52, 62]. The chosen cut-off values, however, varied substantially.

MRI characteristics are outlined in Online Resource 4. Magnetic field strength for all included papers was 1.5 Tesla. Sequences used were T1-weighted dual spin echo, T1-weighted dual gradient echo or T1-weighted spoiled gradient echo for in-phase and out-of-phase (IP/OP) chemical shift imaging. Two studies also evaluated T2-weighted fast spin
echo imaging with and without fat suppression (±FS). Liver steatosis was evaluated by the amount of signal intensity (SI) loss on OP images compared with IP images and by SI difference between FS and non-FS images. Exact measuring methods, however, differed. Correction for T2* effects was performed by d’Assignies et al [24]. Cho et al were the only authors to define cut-off values prospectively [22].

1H-MRS imaging characteristics are outlined in Online Resource 5. Magnetic field strengths were 1.5 T(3/5) and 3 T (2/5). Four papers used a point-resolved spectroscopic sequence (PRESS), one in combination with chemical-shift selective water suppression (CHESS). Krššák et al used a stimulated echo acquisition mode (STEAM) sequence [41]. Voxel sizes varied from 18×18×18 mm to 30×30×40 mm. Hepatic steatosis was evaluated by the ratio of lipid versus water peaks and by the choline-to-lipid ratio. One paper each corrected for T2 effects or T1 and T2 effects [24, 41]. All included studies defined cut-off values retrospectively.

Reference test Details of the reference test are outlined in Online Resource 6. Cut-off values for grading steatosis severity differed among the articles included, complicating their comparison. Two studies compared semi-quantitative visual analysis of steatosis with the automatic vacuole segmentation method [24] or gas-liquid chromatography [41]. Thirty studies examined the presence of fibrosis, 24 examined the presence of inflammatory activity and eight studies examined the presence of iron.

Data for calculation of diagnostic accuracy Extraction of 2×2 accuracy data resulted in 48 complete data sets for US; 33 for CT; 15 for MRI and 7 for 1H-MRS. For both CT and MRI, 5 datasets were not included for analysis because the examinations were visually assessed. This resulted in 28 datasets being analysed for CT and 10 datasets being analysed for MRI. Datasets with TP and FN only were also included for analysis: 4 for US and 1 for CT. The number of datasets per cut-off group is noted in Table 2. Six authors were contacted for completion or verification of data; three answered of which one supplied additional datasets. Three studies reported data-sets for multiple readers [30, 31, 39].
## Table 1  Patient and design characteristics of 46 papers included

<table>
<thead>
<tr>
<th>Study</th>
<th>Technique</th>
<th>Design</th>
<th>No of patients analysed (m/f)</th>
<th>Patient spectrum (n)</th>
<th>Time interval between reference test and index test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yajima 1983</td>
<td>US</td>
<td>P</td>
<td>45 (ND/ND)</td>
<td>10 fatty liver; 8 chronic/acute hepatitis; 17 cirrhosis; 1 PSC; 9 other</td>
<td>≤14 days</td>
</tr>
<tr>
<td>Korte 1986</td>
<td>US</td>
<td>P</td>
<td>100 (67/33)</td>
<td>Suspected liver disease: alcohol abuse, diabetes, obesity</td>
<td>Mean 1.9 weeks</td>
</tr>
<tr>
<td>Saverymuttu 1986</td>
<td>US</td>
<td>P</td>
<td>85 (ND/ND)</td>
<td>50 ALD; 12 active hepatitis; 2 CP; 7 PBC; 2 cryptogenic cirrhosis; 7 other; 4 normal</td>
<td>≤28 days</td>
</tr>
<tr>
<td>Lossner 1988</td>
<td>US</td>
<td>P</td>
<td>187 (ND/ND)</td>
<td>Suspected liver disease</td>
<td>Mean 11 days</td>
</tr>
<tr>
<td>Joseph 1991</td>
<td>US</td>
<td>P</td>
<td>50 (ND/ND)</td>
<td>30 ALD; 5 chronic persistent/reactive hepatitis; 4 cryptogenic cirrhosis; 2 PBC; 9 other</td>
<td>0 days</td>
</tr>
<tr>
<td>Caturelli 1992</td>
<td>US</td>
<td>P</td>
<td>35 (21/14)</td>
<td>Suspected liver disease</td>
<td>Unclear</td>
</tr>
<tr>
<td>Hultcrantz 1993</td>
<td>US</td>
<td>P</td>
<td>83 (47/36)</td>
<td>45 fatty liver; 14 cirrhosis; 11 chronic inflammation; 13 other</td>
<td>0 days</td>
</tr>
<tr>
<td>Dietrich 1998</td>
<td>US</td>
<td>P</td>
<td>68 (40/28)</td>
<td>HCV</td>
<td>Unclear</td>
</tr>
<tr>
<td>Kutcher 1998</td>
<td>US</td>
<td>R</td>
<td>64 (39/25)</td>
<td>HCV</td>
<td>Unclear</td>
</tr>
<tr>
<td>Graif 2000</td>
<td>US</td>
<td>P</td>
<td>28 (ND/ND)</td>
<td>12 HCV; 1 HBV; 1 PBC; 3 non alcoholic cirrhosis; 11 other</td>
<td>≤7 days</td>
</tr>
<tr>
<td>Rinella 2001</td>
<td>CT, MRI</td>
<td>R</td>
<td>33 (19/14)</td>
<td>Potential living liver donors</td>
<td>Unclear</td>
</tr>
<tr>
<td>Mathiesen 2002</td>
<td>US</td>
<td>P</td>
<td>165 (110/55)</td>
<td>65 NAFLD; 25 HCV; 14 ALD; 3 NASH; 3 AIH; 2 PBC; 1 A1-antitrypsin deficiency; 52 non-specific</td>
<td>87%≤28 days;</td>
</tr>
<tr>
<td>Saadeh 2002</td>
<td>US, CT, MRI</td>
<td>P</td>
<td>25 (11/14)</td>
<td>8 NAFLD; 17 NASH</td>
<td>≤90 days</td>
</tr>
<tr>
<td>Kichian 2003</td>
<td>US</td>
<td>R</td>
<td>49 (25/24)</td>
<td>NAFLD</td>
<td>Unclear</td>
</tr>
<tr>
<td>Rinella 2003</td>
<td>MRI</td>
<td>ND</td>
<td>22 (13/9)</td>
<td>15 potential living liver donors; 7 NAFLD</td>
<td>Unclear</td>
</tr>
<tr>
<td>Iwasaki 2004</td>
<td>CT</td>
<td>P</td>
<td>266 (137/129)</td>
<td>Living liver donors</td>
<td>Unclear</td>
</tr>
<tr>
<td>Limanond 2004</td>
<td>CT</td>
<td>R</td>
<td>42 (29/13)</td>
<td>Potential living liver donors</td>
<td>≤28 days</td>
</tr>
<tr>
<td>Hepburn 2005</td>
<td>US</td>
<td>R</td>
<td>164 (108/56)</td>
<td>HCV</td>
<td>Mean 1–2 months; range 0–6 months</td>
</tr>
<tr>
<td>Kim SH 2005</td>
<td>US</td>
<td>R</td>
<td>94 (71/23)</td>
<td>Potential living liver donors</td>
<td>Mean 39.9 days; range 0–140 days</td>
</tr>
<tr>
<td>Kim SH 2006</td>
<td>MRI</td>
<td>R</td>
<td>57 (40/17)</td>
<td>Potential living liver donors</td>
<td>Mean 14.5 days ± 18.8 days; range 0–124 d</td>
</tr>
<tr>
<td>Palmentieri 2006</td>
<td>US</td>
<td>P</td>
<td>235 (127/108)</td>
<td>146 HCV; 30 HBV; 3 HCV/HBV; 33 NAFLD; 23 other</td>
<td>Unclear</td>
</tr>
<tr>
<td>Park 2006</td>
<td>CT</td>
<td>P</td>
<td>154 (104/50)</td>
<td>Potential living liver donors</td>
<td>0 days</td>
</tr>
<tr>
<td>Hamaguchi 2007</td>
<td>US</td>
<td>P</td>
<td>94 (ND/ND)</td>
<td>68 NAFLD; 26 no disease</td>
<td>≤30 days</td>
</tr>
<tr>
<td>Hirsch 2007</td>
<td>US</td>
<td>P</td>
<td>122 (78/44)</td>
<td>HCV</td>
<td>Unclear</td>
</tr>
<tr>
<td>Lee JY 2007</td>
<td>US, CT</td>
<td>R</td>
<td>589 (408/181)</td>
<td>Potential living liver donors</td>
<td>Unclear</td>
</tr>
<tr>
<td>Lee SW 2007</td>
<td>CT</td>
<td>R</td>
<td>48 (44/4)</td>
<td>Living liver donors</td>
<td>0 days</td>
</tr>
<tr>
<td>Perez 2007</td>
<td>US</td>
<td>R</td>
<td>131 (78/53)</td>
<td>116 HCV; 5 HBV; 10 other</td>
<td>≤9 months</td>
</tr>
<tr>
<td>Bahl 2008</td>
<td>MRI</td>
<td>P</td>
<td>52 (15/37)</td>
<td>29 NAFLD; 23 HCV/HHV</td>
<td>≤4 months</td>
</tr>
<tr>
<td>Chen 2008</td>
<td>US</td>
<td>R</td>
<td>108 (75/33)</td>
<td>108 HCV</td>
<td>Unclear</td>
</tr>
<tr>
<td>Cho 2008</td>
<td>MRI, CT</td>
<td>R</td>
<td>131 (66/65)</td>
<td>Liver resection for benign (5) or malignant disease (126)</td>
<td>Median 17 days</td>
</tr>
<tr>
<td>Orlacchio 2008</td>
<td>¹H-MRS</td>
<td>P</td>
<td>30 (17/13)</td>
<td>HCV</td>
<td>≤30 days</td>
</tr>
<tr>
<td>Yoshimitsu 2008</td>
<td>CT, MRI</td>
<td>R</td>
<td>58 (35/23)</td>
<td>38 potential living liver donors; 20 liver metastases</td>
<td>≤14 days</td>
</tr>
<tr>
<td>Crum-Cianflone 2009</td>
<td>US</td>
<td>P</td>
<td>216 (204/12)</td>
<td>HIV</td>
<td>Mean 5.2 months ± 3.6 months</td>
</tr>
<tr>
<td>Dasarathy 2009</td>
<td>US</td>
<td>P</td>
<td>73 (48/25)</td>
<td>38 HCV; 21 NAFLD; 7 HBV; 7 other</td>
<td>0 days</td>
</tr>
<tr>
<td>d’Assignies 2009</td>
<td>MRI, ¹H-MRS</td>
<td>P</td>
<td>20 (15/5)</td>
<td>14 NAFLD; 6 ALD</td>
<td>≤60 days</td>
</tr>
</tbody>
</table>
Group 1 (cut-off values of >0%, >2% and >5% steatosis) Sensitivity and specificity estimates were 73.3% and 84.4% for US; 46.1% and 93.5% for CT; 82.0% and 89.9% for MRI and 88.5% and 92.0% for $^1$H-MRS, respectively. The sensitivity of $^1$H-MRS was significantly higher than that of US ($p<0.05$) and CT ($p<0.01$) and the...
Table 2 Summary estimates of US, CT, MRI and 1H-MRS per combined cut-off value group

<table>
<thead>
<tr>
<th>Group</th>
<th>&gt;0%, &gt;2%, &gt;5% steatosis as positives on biopsya</th>
<th>p-value</th>
<th>n a</th>
<th>Sensitivity (95% CI)</th>
<th>US</th>
<th>CT</th>
<th>MRI</th>
<th>n a</th>
<th>Specificity (95% CI)</th>
<th>US</th>
<th>CT</th>
<th>MRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>US 19</td>
<td>73.3 (62.2–82.1)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>CT 6</td>
<td>46.1 (22.2–71.8)</td>
<td>NS</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>MRI 5</td>
<td>82.0 (63.7–92.2)</td>
<td>NS</td>
<td>0.02</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>1H-MRS 3</td>
<td>88.5 (76.6–94.7)</td>
<td>0.04</td>
<td>&lt;0.01</td>
<td>NS</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Group 2: &gt;10%, &gt;15%, &gt;20% steatosis as positives on biopsyb</td>
<td>p-value</td>
<td>n a</td>
<td>ln DOR (95% CI)</td>
<td>US</td>
<td>CT</td>
<td>MRI</td>
<td>n a</td>
<td>Sensitivity (95% CI)</td>
<td>US</td>
<td>CT</td>
<td>MRI</td>
<td>n a</td>
</tr>
<tr>
<td>US 19</td>
<td>2.70 (1.97–3.43)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>CT 6</td>
<td>2.50 (1.13–3.87)</td>
<td>NS</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>MRI 5</td>
<td>3.70 (2.50–4.90)</td>
<td>NS</td>
<td>NS</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>1H-MRS 3</td>
<td>4.48 (3.15–5.81)</td>
<td>0.02</td>
<td>0.04</td>
<td>NS</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Group 3: &gt;25%, &gt;30%, &gt;33% steatosis as positives on biopsyc</td>
<td>p-value</td>
<td>n a</td>
<td>ln DOR (95% CI)</td>
<td>US</td>
<td>CT</td>
<td>MRI</td>
<td>n a</td>
<td>Sensitivity (95% CI)</td>
<td>US</td>
<td>CT</td>
<td>MRI</td>
<td>n a</td>
</tr>
<tr>
<td>US 19</td>
<td>85.7 (78.4–90.8)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>CT 15</td>
<td>72.0 (59.7–81.7)</td>
<td>0.03</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>MRI 3</td>
<td>97.4 (83.5–99.6)</td>
<td>NS</td>
<td>0.01</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>1H-MRS 2</td>
<td>72.7 (41.4–91.0)</td>
<td>NS</td>
<td>NS</td>
<td>0.03</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Group 4: &gt;50%, &gt;60%, &gt;66% steatosis as positives on biopsyd</td>
<td>p-value</td>
<td>n a</td>
<td>ln DOR (95% CI)</td>
<td>US</td>
<td>CT</td>
<td>MRI</td>
<td>n a</td>
<td>Sensitivity (95% CI)</td>
<td>US</td>
<td>CT</td>
<td>MRI</td>
<td>n a</td>
</tr>
<tr>
<td>US 9</td>
<td>4.75 (3.41–6.80)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>CT 0</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

sensitivity of MRI was significantly higher than that of CT (p=0.02). No significant differences in specificity were found. The lnDOR of 1H-MRS was significantly higher compared with US (p=0.02) and CT (p=0.04).

Group 2 (cut-off values of >10%, >15% and >20% steatosis) Sensitivity and specificity estimates were 90.5% and 69.6% for US; 57.0% and 88.1% for CT; 90.0% and 95.3% for MRI and 82.6% and 94.3% for 1H-MRS, respectively. CT had a significantly lower sensitivity compared with US, MRI and 1H-MRS (p<0.01). Although US had a sensitivity comparable to MRI and 1H-MRS, the specificity was significantly lower than CT (p<0.01), MRI (p<0.01) and 1H-MRS (p=0.01). The lnDOR of MRI was significantly higher than the lnDOR for both US (p=0.05) and CT (p<0.01). 1H-MRS had a significantly higher lnDOR than CT (p=0.03).

Group 3 (cut-off values of >25%, >30% and >33% steatosis) Sensitivity and specificity estimates were 85.7% and 85.2% for US; 72.0% and 94.6% for CT; 97.4% and 76.1% for MRI and 72.7% and 95.7% for 1H-MRS, respectively. The sensitivity of MRI was significantly higher than CT (p=0.01) and 1H-MRS (p=0.03). The specificity of

---

*Number of datasets analysed

*For CT and MRI: Only data for >0% and >5% available

*For US and 1H-MRS: Only data for >10% steatosis available

*CT includes data for >25% and >30% steatosis only; MRI includes data for >30% only; 1H-MRS includes data for >30% and >33% only

**ln DOR = logit sensitivity + logit specificity. NS = not significant

---
MRI however was significantly lower than both CT \( (p=0.02) \) and \(^1\)H-MRS \( (p=0.04) \). Further, the sensitivity for US was significantly higher than for CT \( (p=0.03) \), the specificity for US was significantly lower than for CT \( (p=0.03) \). Analysis of the lnDOR did not show any significant differences.

**Group 4 (cut-off values of >50%, >60% and >66% steatosis)** For this group, data analysis was possible for US only. Sensitivity and specificity estimates were 91.1% and 91.9% respectively.

Figure 3 shows the diagnostic performances (lnDOR) of all imaging techniques per cut-off value group, illustrating the better performance for both MRI and \(^1\)H-MRS compared with US and CT.

**Discussion**

Our results show that MRI and \(^1\)H-MRS perform better than US and CT over the total range of cut-off values that were analysed. For the lower cut-off ranges, we found significant differences in favour of both MRI and \(^1\)H-MRS.

These findings suggest that MRI and \(^1\)H-MRS also perform better than US and CT for detecting separate disease grades, especially for mild disease (<30% steatosis). This is of value in clinical practice when an accurate estimation of the amount of hepatic steatosis is needed. Additional benefits of MRI and \(^1\)H-MRS over US are the quantitative measurements which are less subject to inter- and intraobserver variability [64]. For CT, drawbacks are the radiation exposure and factors affecting the accuracy of the results, such as imaging parameters or iron accumulation [11, 65].

Several limitations of our study must be considered. First, the studies included showed great heterogeneity regarding patient spectrum, reference test, index test and data reporting. Therefore, comparison of separate disease grades and sub-analysis of different aetiologies of steatosis (e.g. NALFD/NASH versus HCV) was precluded. Standardisation of future study designs is needed to enable these comparisons. Moreover, no studies compared all four imaging techniques within the same population, which would be the ideal study design. We were therefore restricted to summarising accuracy data for each technique separately across all the studies included. These indirect comparisons of studies, which showed substantial methodological heterogeneity, might have biased our results.

Second, we had to make the decision to group accuracy results from different cut-off values into four subgroups to enable meta-analysis and to reduce the number of summary estimates and comparisons. The ideal situation would have been to analyse accuracy results for each cut-off value separately.

Third, a standard method for meta-analysis of diagnostic studies is the summary Receiver Operating Characteristic (sROC). For the sROC approach, a negative correlation between the logit sensitivity and the logit specificity is required [17]. As we did not find this negative correlation in our data, plotting of sROC curves was not possible. We therefore used the lnDOR to summarise our results.

Fourth, we did not analyse 3×3 or 4×4 data as the reporting thereof was scarce. By dichotomising the results, we lost information on the capability of imaging techniques to diagnose the degree of steatosis.

**Fig. 3** Comparison of logarithmic diagnostic odds ratios of US, CT, MRI and \(^1\)H-MRS
A fifth limitation was that we chose to exclude articles with 1H-MRS as the reference standard [66–73]. 1H-MRS is increasingly used as a reference standard for steatosis quantification since the results from the Dallas Heart Study were published by Szczepaniak et al in 2005 [74]. However, no clear consensus on this topic currently exists. The articles that were excluded all compared MRI with 1H-MRS and showed good correlations. Therefore, only a small number of datasets were available for analysis of MRI. Additionally, the included articles for MRI did not evaluate triple-echo, multi-echo or multi-interference techniques, whereas the aforementioned excluded articles did. Guiu et al recently suggested that these new techniques should replace the classical dual-echo chemical shift imaging methods, which are not reliable for quantification of liver fat in the case of liver iron overload because of T2* effects [75]. We believe that the small number of available data in combination with the techniques used could have negatively influenced our accuracy results for MRI.

We therefore recommend that consensus on the role of 1H-MRS as the reference standard needs to be established. For liver biopsy evaluation, we recommend using the classification from Kleiner et al for a uniform grading of hepatic steatosis [76].

In conclusion, we have shown that MRI and 1H-MRS are most accurate for the detection of hepatic steatosis. For future research, it is important to improve the study design and reporting of accuracy results.

Acknowledgements J.G. Daams (clinical librarian) for the electronic database search. L. Alvarez-Herrero, S.J. Goncalves, C. Lavini of Academic Medical Center, University of Amsterdam, The Netherlands and T. Takahara of University Medical Center Utrecht, The Netherlands for translating articles.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

Noninvasive quantitation of human liver steatosis using magnetic resonance and bioassay methods. Eur Radiol 19:2033–2040
imaging in nonalcoholic fatty liver disease. Gastroenterology 123:745–750