Quantitative imaging of liver fat and fibrosis
Bohte, A.E.

Citation for published version (APA):
Bohte, A. E. (2014). Quantitative imaging of liver fat and fibrosis

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
ULTRASOUND CANNOT PREDICT THE PRESENCE OR SEVERITY OF HEPATIC STEATOSIS IN SEVERELY OBESE ADOLESCENTS


Published in Radiology 2012; 262(1): 327-34
ABSTRACT

Purpose
To evaluate the diagnostic accuracy of ultrasound (US) for the assessment of hepatic steatosis in severely obese adolescents with proton-MR-spectroscopy (1H-MRS) as the reference standard, and to provide insight on the influence of prevalence on predictive values by calculating positive and negative post-test probabilities.

Materials and methods
This prospective study was IRB approved. All participants, and/or their legal representatives, gave written informed consent. Sensitivity, specificity, positive and negative predictive values (PPV, NPV) were calculated for the overall presence of steatosis and for the presence of substantial (moderate/severe) steatosis. Positive and negative post-test probabilities were calculated and plotted against prevalence.

Results
104 children (47 male, 57 female) were prospectively included. Mean age was 14.5 years (range 8.3-18.9) and mean age adjusted standard deviation BMI score (BMI z-score) was 3.3 (range 2.6- 4.1). The overall prevalence of hepatic steatosis was 46.2% (48/104). Sensitivity of US was 85.4% (41/48), specificity was 55.4% (31/56), PPV was 62.1% (41/66) and NPV was 81.6% (31/38). The prevalence of substantial steatosis was 15.4% (16/104), with US sensitivity of 75.0% (12/16) and specificity of 87.5% (77/88). PPV was 52.2% (12/23) and NPV was 95.1% (77/81). Plotting of post-test probabilities against prevalence for both disease degrees demonstrated how disease prevalence influences US accuracy.

Conclusion
Positive US examination results in severely obese adolescents cannot accurately predict the presence and severity of hepatic steatosis and require additional imaging. Negative US results exclude the presence of substantial steatosis with acceptable accuracy. Steatosis prevalence differs among specific populations, strongly influencing post-test probabilities.
INTRODUCTION
Nonalcoholic fatty liver disease (NAFLD) is currently the most common cause of liver disease in youth. Its prevalence increases concomitant with the epidemic of childhood obesity [1]. The clinical spectrum of NAFLD ranges from simple steatosis to nonalcoholic steatohepatitis (NASH), fibrosis and cirrhosis. As in adults, pediatric NAFLD is associated with components of the metabolic syndrome (defined by the presence of obesity, dyslipidemia, insulin resistance and hypertension) and cardiovascular diseases. NAFLD related liver complications have been observed in children, emphasizing the need to detect, monitor and treat NAFLD in childhood [2-4].

Diagnosing NAFLD remains problematic as it formally requires a liver biopsy which is unsuited for screening purposes due to its invasive nature. Non-invasive diagnostic tools, mostly alanine aminotransferase (ALT) levels or ultrasound (US), are therefore used for diagnosing NAFLD in trials and in clinical practice, especially in the pediatric population [5]. Serum ALT, however, correlates somewhat with inflammation, but not with liver steatosis. US on the other hand has an adequate diagnostic accuracy to detect moderate/severe hepatic steatosis in the adult population in general, with pooled sensitivities ranging from 85.7–91.1% and pooled specificities ranging from 85.2–91.9% in a recently published meta-analysis [6]. All data were obtained from studies that compared US with histopathology. To date, no such data exist for the pediatric population. Proton magnetic resonance imaging (1H-MRS) has gained ground over the past decade as an alternative and noninvasive reference standard for liver fat content evaluation because of its high diagnostic accuracy and reproducibility [7, 8].

To the best of our knowledge, two studies have been published reporting the accuracy of US for steatosis evaluation in the obese pediatric population, both using dual-echo MRI as reference standard [9, 10]. Dual-echo MRI, however, is not an optimal reference standard owing to the technique’s susceptibility to confounders [11]. Moreover, predictive values were not reported. When clinicians want to determine whether a specific test result reflects the correct diagnosis, predictive values — expressed as positive and negative post-test probabilities — are more informative than sensitivity or specificity, provided that disease prevalence is considered [12].

The purpose of this study was therefore (1) to evaluate the diagnostic accuracy of US for the assessment of hepatic steatosis in a cohort of severely obese adolescents with proton MR-spectroscopy (1H-MRS) as the reference standard, and (2) to provide a clear insight on the influence of prevalence on predictive values by calculating the positive and negative post-test probabilities, based on the obtained diagnostic accuracy estimates.

MATERIAL AND METHODS
Study Design and Patients
This study was approved by the institutional review board of the Academic Medical Center Amsterdam, The Netherlands. Written informed consent was obtained from parents or legal representatives of participants who were under 12 years of age. For those between 12 and 18 years of age, informed consent was obtained from both the participant and the parents or legal representatives. Between February 2008 and October 2010, we prospectively included a cohort of 121 consecutive severely obese children who were admitted to a tertiary center-based lifestyle intervention program. Admission criteria for this program were: severe primary obesity...
(body mass index adjusted for age >35 kg/m²) or primary obesity (BMI-for-age >30 kg/m²) along with obesity-related co-morbidity. BMI-for-age was reported as BMI z-score, which reflects the number of standard deviations (SD) from the mean on a standard BMI curve for age and gender. Children with a BMI z-score of >2 (95th percentile) are considered obese, those with a BMI z-score of >2.6 (99th percentile) are considered severely obese [13, 14]. All admitted children between 8-18 years old were eligible for participation in our study. Exclusion criteria were: known liver disease other than NAFLD (viral/autoimmune hepatitis, Wilson disease, hemochromatosis, α1-antitrypsin deficiency); metabolic disease (β-oxidation defects, urea cycle defects); (history of) use of steatogenic medication; alcohol consumption >7 units/week; jejuno-ileal surgery; history of parenteral feeding and contra-indications for MRI. Before participating in the weight-loss program, all children underwent a liver ultrasound examination and a ¹H-MRS scan of the liver in a tertiary care setting within a time interval of one month.

**Imaging Techniques**

**Index Test: Ultrasound**

US examinations were performed and interpreted by one of three radiologists (T.P.R, S.A.S. and E.M.K.), 5-20 years experience, >600 liver US examinations per year). All were blinded to clinical and ¹H-MRS data. Philips ATL HDI 5000, HD11 and IU22 US systems were used with 2-5 and 3.75 MHz curved array transducers. The US examinations for this study were incorporated in the daily workload of a teaching hospital and reflect routine clinical practice. The “abdominal general” setting was used on all US systems. Gain and focus were manually adjusted, depending on patient habitus and beam attenuation. For the diagnosis and grading of severity of hepatic steatosis, the following standardized views of the liver were obtained: transverse and longitudinal views of the right hepatic lobe, including the right kidney and diaphragm; a sagittal view of the left liver lobe; a view including the portal vasculature, and a view through the gall bladder region. On these images, the following four widely accepted scoring items were evaluated [15]: (1) echogenicity of liver parenchyma; (2) visualization of diaphragm; (3) visualization of intrahepatic vessels and (4) visualization of posterior part of the right hepatic lobe.. A final qualitative score from 0 to 3 was given with respect to liver steatosis (Table 1, Figure 1): score 0 for no steatosis, score 1 for mild steatosis, score 2 for moderate steatosis and score 3 for severe steatosis.

**Interobserver and intraobserver agreement**

Nineteen consecutive patients were asked for permission to have extra US examinations performed for the analysis of inter and intraobserver agreement. All three radiologists performed US liver examinations in this patient subgroup on the same day for interobserver agreement analysis. For intraobserver agreement analysis, US examinations were repeated one week later in 16 of 19 patients by the radiologists. Three patients did not have the second US examination performed due to logistic problems.

**Reference Test: Proton MR- Spectroscopy**

¹H-MR spectra were acquired by using first order iterative shimming and a point-resolved spectroscopy sequence (TE/TR=38/2000 ms, 64 signal acquisitions) in a voxel of 20 x 20 x 20 mm
Figure 1. Examples of liver ultrasound examinations: none, mild, moderate and severe steatosis. A: No steatosis (score 0). Longitudinal view of liver and right kidney of a 14 year old boy, BMI-z score 3.3. Ultrasound examination shows normal liver parenchyma echogenicity (LP), comparable to that of the kidney parenchyma (K). Normal visibility of diaphragm (D) and intrahepatic vessels (V). Proton magnetic resonance spectroscopy (1H-MRS): 0.4% liver fat concentration (LFC). B: Mild steatosis (score 1): 11 year old boy; BMI-z score: 3.0. Increased echogenicity of liver parenchyma compared with kidney parenchyma. Normal visualization of diaphragm and slightly impaired visualization of intrahepatic vessels. Corresponding 1H-MRS result: 6.0% LFC. C: Moderate steatosis (score 2): 12 year old boy; BMI-z score: 3.4. Increased echogenicity of liver parenchyma compared with kidney parenchyma. Impaired visualization of diaphragm and intrahepatic vessels. 1H-MRS: 14.8% LFC. D: Severe steatosis (score 3): 13 year old boy (BMI-z score: 3.9). Ultrasound examination shows no visibility of diaphragm or intrahepatic vessels. The radiologist was unable to visualize the kidney. Ultrasound examination overestimated disease degree: LFC with 1H-MRS was 4.2% (indicating mild steatosis). BMI-z: body mass index z-score (standard deviation of mean BMI adjusted for age).

Table 1. Scoring of hepatic steatosis with Ultrasound

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (no steatosis)</td>
<td>Normal echogenicity of liver parenchyma, normal visualization of diaphragm and intrahepatic blood vessels</td>
</tr>
<tr>
<td>1 (mild steatosis)</td>
<td>Slightly increased echogenicity of liver parenchyma, normal visualization of diaphragm and intrahepatic blood vessels</td>
</tr>
<tr>
<td>2 (moderate steatosis)</td>
<td>Markedly increased echogenicity of liver parenchyma, slightly impaired visualization of diaphragm and intrahepatic vessels</td>
</tr>
<tr>
<td>3 (severe steatosis)</td>
<td>Severely increased echogenicity of liver parenchyma with poor or no visualization of diaphragm and intrahepatic vessels and posterior part of the right liver lobe</td>
</tr>
</tbody>
</table>
on a 3.0 T MR scanner during free breathing (Intera, Philips Healthcare, Best, The Netherlands) (Figure 2 a, b). If patients weighted over 150 kilograms or if they had an abdominal circumference of more than 150 cm, the MR scans were performed on an open bore 1.0 T MR scanner (Panorama, Philips Healthcare). Six-channel (3T) and three-channel (1T) torso coils were used. The fat (methylene, CH$_2$) and water resonance peaks, located at 1.3 ppm and 4.7 ppm respectively (Figure 2), were fitted with the AMARES algorithm of the spectroscopic signal processing package jMRUI [16]. Calculated peak areas were corrected for average T2 relaxation effects using the following T2 values for 3T: $T2_{\text{water}3T} = 34$ ms, $T2_{\text{fat}3T} = 68$ ms [17]. For 1T, T2 values of $T2_{\text{water}1T} = 69$ ms and $T2_{\text{fat}1T} = 60$ ms were used [18]. The relative fat signal fraction (FSF) was calculated as:

$$FSF = \frac{\text{fat}}{\text{fat} + \text{water}}$$

FSF was further converted to the fat volume fraction (FVF):

$$FVF = \frac{FSF}{1.138 - 0.339 \cdot FSF}$$

Subsequently, the absolute mass concentration of liver fat (LFC) was calculated using a liver tissue density of 1.051 g/L and a density of fat in the liver of 0.90 g/L:

---

Figure 2 A,B. Example of proton magnetic resonance spectroscopy ($^1$H-MRS) voxel positioning in (A) coronal and (B) transverse plane of T2-weighted anatomical liver MR series (girl, 11 years old). Large vessels and liver edges have been avoided for voxel placement.

Figure 2 C-E. Examples of $^1$H-MR spectra obtained in a (C) 13 year old girl without hepatic steatosis, (D) 11 year old girl with mild steatosis (4% LFC) and (E) 17 year old boy with severe hepatic steatosis (24% LFC). The water peak is visible at a frequency of 4.7 ppm in all three examples. The fat (methylene) peak at 1.3 ppm is visible in (D) and (E).

LFC: liver fat concentration; ppm: parts per million; a.u.: arbitrary units.
LFC (mg/g wet weight) = FVF · 0.856 [19, 20].

Presence of hepatic steatosis was defined as >1.8% LFC measured with 1H-MRS. Presence of substantial (moderate/severe) hepatic steatosis was defined as >7.7% LFC. Both cutoffs have been validated to correspond with >5% and >33% fat containing hepatocytes on liver biopsy, respectively [21, 22]. 1H-MR spectra were analyzed by a research fellow (A.E.B, 3 years of experience) under supervision of an MR physicist (A.J.N, 8 years of experience), who were both blinded to the results of the US liver examination.

**Statistical Analysis**

**Demographic data**

Differences in age, BMI-z score and LFC between boys and girls were calculated with the independent sample t-test for parametric data and Mann Whitney U test for non-parametric data. A p-value <0.05 was considered statistically significant.

**Diagnostic accuracy**

Accuracy, underestimation and overestimation of US compared to 1H-MRS were calculated. For both thresholds on 1H-MRS, sensitivity, specificity, positive and negative predictive values (PPV, NPV) and positive and negative likelihood ratios (LR+, LR-) were calculated with 95% confidence intervals (CI).

**Interobserver and intraobserver agreement**

For assessment of interobserver and intraobserver agreement, linear weighted kappa scores with 95% confidence intervals were calculated. A kappa score below 0.4 indicated poor agreement, a score of 0.4-0.6 moderate, 0.6-0.8 good and 0.8-1.0 excellent agreement [23].

**Post-test probabilities related to prevalence**

Positive and negative post-test probabilities were calculated based on the Bayesian principles, using positive and negative LR and pretest odds [12]. As the prevalence of hepatic steatosis differs, depending on various risk factors, we plotted positive and negative post-test probabilities of US against a prevalence range of 0-100%.

All analyses were performed with PASW Statistics 18; SPSS inc., Chicago, IL and with Microsoft Office Excel; Microsoft; Redmond; WA.

**RESULTS**

**Patients**

We included 121 children in this study. Two children withdrew from the study before study procedures were finished and data of US and 1H-MRS were incomplete in six children. The time interval between US and 1H-MRS exceeded one month for nine children, who were also excluded from analysis. In total, data of 104 children (47 male, 57 female) were analyzed. The mean age of these 104 children was 14.5 ± 2.2 years (range: 8.3-18.9). Mean BMI-z score was 3.3 ± 0.3 (range: 2.6-4.1). There was no statistically significant age difference between boys and girls (p=0.49); BMI-z score of boys was significantly higher than that of girls (p=0.002), **Table 2**.
Imaging results

$^1$H-MRS was performed at 3T MRI for 89 children, and at open 1T MRI for 15 children. The prevalence of hepatic steatosis was 46.2% (48 of 104). Substantial steatosis was present in 15.4% (16 of 104) of children. Median LFC was 1.7% (range: 0.2 – 26.0%). There was no statistically significant difference in LFC between boys and girls ($p = 0.13$), Table 2.

US results are illustrated in Figure 3. US diagnoses were correct for 57.7% of patients (60 of 104). US examination overestimated disease severity in 33.7% of patients (35 of 104) (Table 3). Disease severity was underestimated by US for 8.7% of patients (9 of 104).

Diagnostic Accuracy

Overall hepatic steatosis

For detection of overall hepatic steatosis ($^1$H-MRS > 1.8% LFC), sensitivity of US was 85.4% (41 of 48) and specificity was 55.4% (31 of 56). PPV was 62.1% (41 of 66) and NPV was 81.6% (31 of 38). LR+ was 1.91 and LR− was 0.26 (Table 4).

Substantial hepatic steatosis

For detection of substantial steatosis ($^1$H-MRS >7.7% LFC), sensitivity of US was 75.0% (12 of 16) and specificity was 87.5% (77 of 88). PPV was 52.2% (12 of 23) and NPV was 95.1% (77 of 81). The LR+ and LR− were 6.00 and 0.29 respectively (Table 4).

Table 2. Patient characteristics

<table>
<thead>
<tr>
<th></th>
<th>Total (n=104)</th>
<th>Boys (n=47)</th>
<th>Girls (n=57)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td>14.5 ± 2.2 (8.3-18.9)</td>
<td>14.3 ± 2.1 (10.1 - 18.0)</td>
<td>14.6 ± 2.3 (8.3 – 18.9)</td>
<td>0.49</td>
</tr>
<tr>
<td>BMI-z score</td>
<td>3.3 ± 0.3 (2.6-4.1)</td>
<td>3.5 ± 0.3 (2.8 -4.1)</td>
<td>3.2 ± 0.3 (2.6 –4.0)</td>
<td>0.002</td>
</tr>
<tr>
<td>LFC (%)</td>
<td>1.7 (0.2 – 26.0)</td>
<td>2.3 (0.2 – 26.0)</td>
<td>1.4 (0.2 – 11.4)</td>
<td>0.13</td>
</tr>
</tbody>
</table>

$^1$H-MRS: proton magnetic resonance spectroscopy; LFC: liver fat concentration measured with $^1$H-MRS. Data are presented in mean ± SD (range) for age and BMI-z and in median (range) for LFC.

Table 3. Ultrasound imaging scores compared with $^1$H-MRS categorization

<table>
<thead>
<tr>
<th></th>
<th>None (0 – 1.8% LFC)</th>
<th>Mild (1.8 – 7.7% LFC)</th>
<th>Substantial (&gt;7.7% LFC)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>US score 0 (none)</td>
<td>31†</td>
<td>5‡</td>
<td>2‡</td>
<td>38</td>
</tr>
<tr>
<td>US score 1 (mild)</td>
<td>24§</td>
<td>17†</td>
<td>2†</td>
<td>43</td>
</tr>
<tr>
<td>US score 2 (moderate)†</td>
<td>1§</td>
<td>9 §</td>
<td>11 §</td>
<td>21</td>
</tr>
<tr>
<td>US score 3 (severe)†</td>
<td>0 §</td>
<td>1 §</td>
<td>1 §</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>56</td>
<td>32</td>
<td>16</td>
<td>104</td>
</tr>
</tbody>
</table>

$^1$H-MRS: proton magnetic resonance spectroscopy; LFC: liver fat concentration measured with $^1$H-MRS.

* US scores 2 and 3 match $^1$H-MRS category of “substantial” steatosis.
† Correct diagnosis of disease degree by US
‡ Underestimation of disease degree by US
§ Overestimation of disease degree by US
Interobserver and intraobserver agreement

Interobserver agreement was moderate to good, ranging from 0.58 to 0.68 between radiologists. Intraobserver agreement for each of the three radiologists was excellent, ranging from 0.82 to 0.91 (Table 5).

Post-test probabilities related to prevalence

For the observed disease prevalence of 46.2% in our population, the positive post-test probability (or PPV) was 62.1%. The negative post-test probability (or 1-NPV) was 18.4% (Figure 4a). For substantial hepatic steatosis; the observed disease prevalence of 15.4% in our population resulted in a positive post-test probability (or PPV) of 52.2% and a negative post-test probability (or 1-NPV) of 4.9% (Figure 4b).

DISCUSSION

We have shown in our cohort of severely obese children that the value of US is limited with respect to predicting the presence and severity of hepatic steatosis. However, US has acceptable accuracy to predict the absence of substantial hepatic steatosis. This is different from the severely obese adult population, where the prevalence of hepatic steatosis is reported to be higher [24, 25]. Our calculations of post-test probabilities illustrate how predictive values depend strongly on disease prevalence.

Figure 3. Liver fat concentration (LFC) measured in each ultrasound score group by proton magnetic resonance spectroscopy (1H-MRS). The horizontal lines represent the cutoff values for mild steatosis (1.8 - 7.7% LFC) and substantial steatosis (>7.7% LFC). Median LFC per US degree are: 1.1% (range 0.2-10.1) for degree 0; 1.6% (0.4-11.4%) for degree 1; 7.7 (0.5-26.0%) for degree 2 and 9.8% (4.2-15.3%) for degree 3.
Pozatto et al and Pacifico et al compared the accuracy of US with dual-echo chemical shift MRI in severely obese adolescent populations and reported prevalences of hepatic steatosis of 23.3% and 40.0%, respectively [9, 10]. From the reported data, we were able to calculate predictive values: PPV were 56% and 48%, NPV were 94% and 97%, respectively. Interestingly, two studies that compared the accuracy of US with histopathology results in severely obese adults (with mean BMI scores of 43.8 and 47.5 kg/m²), reported much higher prevalences of hepatic steatosis: 89.5% and 91.4% [24, 25]. Again, we calculated predictive values from the reported data. PPV were high: 98.4% and 95.4% respectively. NPV were consequently low: 23.3% and 12.1%, respectively. Additionally, we found one study that compared US accuracy for hepatic steatosis with liver biopsy results in healthy, lean, adult potential living liver donors [26]. The prevalence of hepatic steatosis in this cohort was 11.5% (with a cutoff of 10% fat containing hepatocytes on liver biopsy), PPV was 30.4% and NPV was 96.4%. The reported data from all these studies, obtained in different populations with different disease prevalences, fit adequately within our post-test probability models.

This finding emphasizes that for the individual patient, factors that influence the prevalence of hepatic steatosis (e.g., age, race, presence of components of the metabolic syndrome) have to be taken into consideration as these strongly influence the predictive value of US examination results [5, 27].

This study has some limitations. Firstly, US examinations were performed by three different radiologists. Interobserver variability is a well-known drawback of US examination. In our cohort, intraobserver agreement was excellent for all three radiologists; interobserver agreement was moderate to good. Accuracy results therefore could have been better, had one observer preformed all US examinations. For logistic reasons, this was not possible in our

### Table 4. Ultrasound accuracy data for diagnosing overall steatosis and substantial steatosis.

<table>
<thead>
<tr>
<th>Ultrasound score</th>
<th>Cutoff (LFC %)</th>
<th>Prevalence (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
<th>LR+</th>
<th>LR-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score ≥ 1</td>
<td>1.8</td>
<td>46.2</td>
<td>85.4</td>
<td>55.4</td>
<td>62.1</td>
<td>81.6</td>
<td>1.91</td>
<td>0.26</td>
</tr>
<tr>
<td>(overall steatosis)</td>
<td></td>
<td></td>
<td>(75.4, 95.4)</td>
<td>(42.3, 68.4)</td>
<td>(50.4, 73.8)</td>
<td>(69.3, 93.9)</td>
<td>(1.40, 2.62)</td>
<td>(0.13, 0.54)</td>
</tr>
<tr>
<td>Score ≥ 2</td>
<td>7.7</td>
<td>15.4</td>
<td>75.0</td>
<td>87.5</td>
<td>52.2</td>
<td>95.1</td>
<td>6.00</td>
<td>0.29</td>
</tr>
<tr>
<td>(substantial steatosis)</td>
<td></td>
<td></td>
<td>(53.8, 96.2)</td>
<td>(80.6, 94.4)</td>
<td>(31.8, 72.6)</td>
<td>(90.3, 99.8)</td>
<td>(3.22, 11.6)</td>
<td>(0.12, 0.67)</td>
</tr>
</tbody>
</table>

Data between parentheses are 95% confidence intervals; LFC: liver fat concentration measured with proton magnetic resonance spectroscopy; PPV: positive predictive value; NPV: negative predictive value LR+: positive likelihood ratio; LR-: negative likelihood ratio.

### Table 5. Intraobserver and interobserver agreement for Ultrasound examinations

<table>
<thead>
<tr>
<th>Intraobserver agreement</th>
<th>Interobserver agreement</th>
<th>n</th>
<th>Linear weighted kappa</th>
<th>n</th>
<th>Linear weighted kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiologist 1</td>
<td>Radiologist 1 versus radiologist 2</td>
<td>16</td>
<td>0.82 (0.63, 1.00)</td>
<td>19</td>
<td>0.68 (0.51, 0.85)</td>
</tr>
<tr>
<td>Radiologist 2</td>
<td>Radiologist 2 versus radiologist 3</td>
<td>16</td>
<td>0.91 (0.80, 1.00)</td>
<td>19</td>
<td>0.58 (0.37, 0.80)</td>
</tr>
<tr>
<td>Radiologist 3</td>
<td>Radiologist 1 versus radiologist 3</td>
<td>16</td>
<td>0.87 (0.69, 1.00)</td>
<td>19</td>
<td>0.59 (0.45, 0.73)</td>
</tr>
</tbody>
</table>

Data between parentheses are 95% confidence intervals.
Figure 4. Post-test probability curves for the prediction of hepatic steatosis in relation to disease prevalence. Black line: positive post-test probability curve. Grey line: negative post-test probability curve. Dotted vertical line: disease prevalence of this study cohort. A: Post-test probability curve for the prediction of any hepatic steatosis (LFC > 1.8%; US score ≥1). The prevalence of hepatic steatosis was 46.2%. The positive post-test probability (PPV) was 62.1%, and negative post-test probability (1-NPV) was 18.4%. B: Post-test probability curve for the prediction of substantial hepatic steatosis (LFC > 7.7%; US score ≥2). Prevalence of substantial hepatic steatosis was 15.4%. The positive post-test probability (PPV) was 52.2%, and negative post-test probability (1-NPV) was 4.9%. LFC: liver fat concentration measured with proton magnetic resonance spectroscopy.
study. Moreover, interobserver variability is intrinsic to the subjective US examination and therefore inevitable in daily clinical practice. Secondly, possible confounders for correct LFC measurement with $^1$H-MRS in our cohort are T2 relaxation effects. T2 values are known to vary between individuals, especially in the presence of iron accumulation. However, no patients with known hemochromatosis were included in this study. We did not perform multiecho $^1$H-MRS acquisitions in our cohort for individual T2 correction. Instead, we corrected for T2 effects using average T2 relaxation values that have been reported in literature [17, 18]. This could have led to small inaccuracies in LFC calculations. We do however not believe that this has significantly affected our accuracy data. Another limitation with respect to $^1$H-MRS is the fact that LFC was measured in a single voxel, while steatosis can be heterogeneously distributed. Multiple voxel measurements or spectroscopic imaging of the whole liver would therefore have been ideal to reduce the risk of sampling error. This, however, is time consuming which makes it impractical in clinical practise. A previous study showed an acceptable coefficient of variation of 14% for LFC measurements in different locations of the liver in patients with fatty liver [8]. Moreover, the liver tissue size of 8000 mm$^3$ that is examined with single voxel $^1$H-MRS is approximately 150 times larger than the amount of liver tissue that is examined at liver biopsy, which already diminishes the sampling error risk compared to the reference standard.

When using US to evaluate hepatic steatosis in obese children and adolescents in clinical practice, we recommend additional MR-imaging ($^1$H-MRS or multi-echo chemical shift MRI) if US examination is positive and further information on the presence and/or severity of steatosis is required. A negative US finding excludes the presence of steatosis in the majority of cases, and presence of substantial steatosis is most unlikely. To improve diagnostics in NAFLD, further effort should be aimed at implementation of standardized $^1$H-MRS or multi-echo chemical shift MRI scanning protocols, enabling comparison of results obtained on different MR systems and at different centers, and simplification of post-processing of acquired MR data.

In conclusion, this study shows that a positive US examination cannot accurately predict the presence or severity of hepatic steatosis in severely obese adolescents; a negative US result predicts the absence of substantial hepatic steatosis with acceptable accuracy. Predictive values depend strongly on disease prevalence and can therefore only be applied to specific populations through calculation of post-test probabilities.

REFERENCES


