Applications of magnetic resonance spectroscopy for noninvasive assessment of hepatic steatosis
van Werven, J.R.

Citation for published version (APA):
Chapter 3: Reproducibility of 3.0T ¹H-MRS in hepatic steatosis

REPRODUCIBILITY OF 3.0 TESLA MAGNETIC RESONANCE SPECTROSCOPY FOR MEASURING HEPATIC FAT CONTENT

Jochem R. van Werven
Johannes M. Hoogduin
Aart J. Nederveen
Andre A. van Vliet
Ewa Wajs
Petra Vandenberk
Erik S. G. Stroes
Jaap Stoker

ABSTRACT

Purpose:
To investigate reproducibility of proton magnetic resonance spectroscopy (1H-MRS) to measure hepatic triglyceride content (HTGC).

Materials and methods:
In 24 subjects, HTGC was evaluated using 1H-MRS at 3.0 Tesla. We studied “between-weeks” reproducibility and reproducibility of 1H-MRS in subjects with fatty liver. We also studied within liver variability and within day reproducibility. Reproducibility was assessed by coefficient of variation (CV), repeatability coefficient (RC) and intraclass correlation coefficient (ICC).

Results:
The CV of between weeks reproducibility was 9.5%, with a RC of 1.3% HTGC (ICC 0.998). The CV in fatty livers was 4.1%, with a RC of 1.3% HTGC (ICC 0.997). Within day CV was 4.5%, with a RC of 0.4% HTGC (ICC 0.999). CV for within liver variability was 14.5%.

Conclusion:
Reproducibility of 1H-MRS to measure HTGC for “between-weeks” measurements and in fatty livers is high, which is important for follow up studies. Within liver variability displays a larger variation, meaning that liver fat is not equally distributed and during consecutive measurements the same voxel position should be used.
INTRODUCTION:

Hepatic steatosis (fatty liver) is present in about one third (20.0% - 33.6%) [1, 2] of the general population in Western countries and is associated with a variety of disorders, including obesity, type 2 diabetes, hepatitis and drug toxicities [3]. Liver biopsy is the reference standard for the assessment of hepatic steatosis. The utility of liver biopsy is limited because of its invasiveness, sampling errors, complications such as bleeding and interobserver variability [4-6]. Noninvasive methods such as ultrasonography, Computed Tomography, Magnetic Resonance Imaging (MRI) and Proton Magnetic Resonance Spectroscopy ($^1$H-MRS) have been used to detect hepatic steatosis and attempts have been made to grade hepatic triglyceride content (HTGC) with these methods [1, 2, 7-9]. Of these imaging techniques MRI and $^1$H-MRS are recognized as the most accurate noninvasive techniques for hepatic fat quantification. These techniques permit the breakdown of the MR signal into water and fat signal components and allow quantification of hepatic fat. To date, there are three clinically used MR imaging techniques for the detection and quantification of hepatic fat including chemical shift imaging, frequency-selective imaging, and $^1$H-MRS. Each technique has important advantages and disadvantages and is increasingly used in diagnosis, treatment and follow-up of fatty liver disease [10-12].

$^1$H-MRS has proven to be a very sensitive noninvasive method to detect hepatic triglyceride content [13] and has shown to correlate with liver biopsy results [14-16]. Noninvasive $^1$H-MRS is also suitable to determine HTGC and follow up patients in clinical trials [17]. Despite the increasing use of $^1$H-MRS in determining hepatic steatosis, there is sparse literature addressing the reproducibility of this technique [18-19]. Knowledge of normal variability in measurements is important, especially when consecutive investigations are planned to follow the course of hepatic steatosis during treatment. Literature exists for “within day” and “within liver” reproducibility, indicating coefficients of variation in the range of 3.6 – 8.5% HTGC [1, 20-21] (within day) and 11.0 – 14.5% HTGC [1, 21] (within liver). However in longitudinal studies knowledge of “between weeks” reproducibility is necessary. To our knowledge only Longo et al. [22] studied between weeks reproducibility in two subjects.
In this study we investigated four aspects of reproducibility of 3.0 Tesla \(^1\)H-MRS to measure hepatic fat content. Primarily this concerned I) “between weeks” reproducibility when measurements are repeated after four weeks and II) reproducibility of \(^1\)H-MRS to measure HTGC in subjects with fatty liver. Secondarily III) we investigated “within liver” variability when measuring HTGC in two different voxels in different parts of the liver. We also investigated IV) “within day” reproducibility of \(^1\)H-MRS when two acquisitions in the liver are made on the same day.

**MATERIALS AND METHODS:**

**Study design**

This study was a nonrandomized pilot study in 24 individuals in total: six healthy subjects, six subjects with familial hypobetalipoproteinemia (FHBL) and 12 obese subjects (body mass index (BMI) over 30 kg/m\(^2\)). We included these different subjects to cover a broad spectrum of hepatic fat content. We choose subjects with FHBL as this condition is associated with fatty liver due to triglyceride accumulation in the liver. Obese subjects were chosen as a positive correlation has been established between increased waist circumference and hepatic fat content [20]. This study was approved by the Medical Ethics Committee. All participants gave written informed consent. The study sponsor had no influence on study design or analysis.

**Volunteers:**

In all 24 subjects hepatic triglyceride content in the liver was evaluated using \(^1\)H-MRS. This cohort comprised 12 males and 12 females. Mean age was 49.1 years (range, 22 – 65). Fifteen out of 24 subjects were obese (BMI > 30.0 kg/m\(^2\)) with mean BMI of 31.2 kg/m\(^2\) (range 21.8 – 41.0). Eleven of 24 subjects had features of the metabolic syndrome defined as having at least three risk factors according to the National Cholesterol Education Program Adult Treatment Panel III definition [23]. One of the risk factors is abdominal obesity. In all 24 subjects \(^1\)H-MRS measurements were performed twice. Both \(^1\)H-MRS scans were performed in fasting condition, in the morning between 8.00 and 10.00 am. The second \(^1\)H-MRS scan was scheduled 4 weeks after the first \(^1\)H-MRS scan. Other standardization or lifestyle control was not performed.
We investigated four aspects of reproducibility. Primarily, we investigated subjects for:
I) "Between weeks" reproducibility; All subjects were scanned twice. The second scan was performed 4 weeks later. In total we studied "between weeks" reproducibility in 24 subjects. For both scans the same voxel position was used.
II) "Between weeks" reproducibility of $^1$H-MRS in fatty livers; only subjects with fatty liver were selected. Szczepaniak et al. [1] defined hepatic steatosis as more than 5.6% HTGC measured by $^1$H-MRS. In a subset of 8 subjects with fatty liver (all six subjects with FHBL and two selected obese subjects) we studied "between weeks" reproducibility. In this subgroup we studied reproducibility of $^1$H-MRS in subjects with hepatic steatosis.
Secondarily we chose to investigate 12 subjects for:
III) "Within liver" variability; during the first visit $^1$H-MRS was performed in two different positions (voxels) in the liver, repeated in the same two voxels after 4 weeks. This was performed in the six healthy subjects and the six subjects with FHBL. We compared both voxels within the 4 weeks time span to study within liver variability. This was done to anticipate for heterogeneity of fat in the liver.
IV) "Within day" reproducibility; 12 obese subjects were scanned twice on the same day. The second scan was performed approximately 4 hours after the first; both scans were performed in fasting conditions.

**MR Spectroscopy**

All measurements were performed on a 3.0 Tesla Philips Intera scanner (Philips Healthcare, Best, the Netherlands) using a cardiac coil. A voxel of 20 x 20 x 20 mm was positioned in the right hepatic lobe, avoiding inclusion of the diaphragm and edges of the liver, but also vascular and biliary structures. When two voxels were used to assess within liver variability, different positions were chosen in the right hepatic lobe. Voxel size and time for acquisition were standardized for all subjects. Spectra were acquired using first order iterative shimming, a PRESS sequence with TE/TR=35/2000 ms and 64 signal acquisitions during free breathing. We evaluated the liver $^1$H-MR spectra by using jMRUI software [24]. A ratio from the $^1$H-MR spectra (Figure 1) was calculated and defined as the methylene peak versus the reference H$_2$O peak. Calculated peak areas of water and fat were corrected for T2 relaxation (T2 water= 34msec, T2 fat= 68msec) [25] and percentage hepatic fat content was calculated according to methods described by Szczepaniak et al. [1]. Room time, including taking subjects in and out of the scanner, positioning of the subjects in the scanner, acquisition of localizers, acquisition of axial-coronal-transversal T2 weighted images for voxel planning and performing $^1$H-MR spectroscopy, was 45 minutes.
Chapter 3: Reproducibility of 3.0T $^1$H-MRS in hepatic steatosis

Figure 1: Example $^1$H-MR spectrum and voxel position in the liver. 1) water peak at 4.65 ppm, 2) methylene peak at 1.3 ppm.

Statistical analysis

$^1$H-MRS analysis: The percentage hepatic fat content was the main endpoint of this study. The reproducibility of the measurements was assessed by means of the Bland-Altman method [26] and repeatability coefficient (RC) which allows calculating the 95% limits of agreement. We also used the Intraclass Correlation Coefficient (ICC), as well as the Coefficient of Variation (CV). We chose to use the Bland-Altman method, because this method is more suited to study reproducibility between different measurements than correlation coefficients. The repeatability coefficient was defined as 1.96 times the standard deviation of the mean difference between two measurements [26]. The CV was investigated because most other literature used this method to study reproducibility of $^1$H-MRS, so we were able to compare results. The CV was calculated by dividing the standard deviation of the mean difference between two measurements divided by the mean HTGC of all measurements. To study differences between groups and scanning visits we used nonparametric tests for related samples (Wilcoxon Signed Rank test). A p-value <0.05 was considered significant. For statistical analysis SPSS (SPSS Inc, Chicago, Illinois, USA) was used.
RESULTS:

When performing this $^1$H-MRS study, we encountered no technical failures, and all $^1$H-MRS measurements were of sufficient quality for analysis.

I) “Between weeks” reproducibility of $^1$H-MRS measurements of HTGC in all subjects:

Mean HTGC in the first $^1$H-MRS measurement was 6.8% and did not differ from the second $^1$H-MRS measurement, 7.0% ($p=0.391$). The CV between both scanning sessions was 9.5%. The RC was 1.3% HTGC (Table 1). In Figures 2 and 3 these results are represented in a scatter plot and Bland-Altman plot to show the limits of agreement between both measurements. The ICC between both scanning sessions was 0.998 ($p<0.001$), indicating that these measurements are reproducible.

Figure 2: Scatter plot of between weeks reproducibility.
II) Reproducibility of $^1$H-MRS in fatty livers:

Hepatic steatosis is defined as more than 5.6% HTGC measured by $^1$H-MRS. In this study we identified eight of in total 24 subjects having more than 5.6% hepatic fat (two obese healthy subjects and six subjects with FHBL). Mean HTGC in the first $^1$H-MRS measurement did not differ (16.7%) from the second $^1$H-MRS measurement, 16.7% (p=0.889). CV was 4.1% and RC 1.3% HTGC. The ICC between both scanning session was 0.997 (p<0.001).

III) “Within liver” variability of $^1$H-MRS:

To measure “within liver” reproducibility of $^1$H-MRS HTGC was assessed in 12 subjects by comparing two voxels positioned in different parts of the right liver lobe (right liver lobe defined as segments IV to VIII by Couinaud), also repeated within 4 weeks. Mean HTGC in the first voxel (9.7%) did not differ from the second voxel, 9.5% (p=0.831). The CV in this group was 14.5%. In Figures 4 and 5 HTGC measured by $^1$H-MRS is represented in a scatter plot and Bland-Altman plot to assess agreement between both measurements in two separate voxels in the liver. The ICC between both scanning sessions was 0.996 (p<0.001).

Figure 3: Bland-Altman plot of between weeks reproducibility.
Chapter 3: Reproducibility of 3.0T $^1$H-MRS in hepatic steatosis

Figure 4: Scatter plot of within liver variability.

Figure 5: Bland-Altman plot of within liver variability.
IV) "Within day" reproducibility of $^1$H-MRS measurements of HTGC:
In total 12 healthy obese subjects underwent $^1$H-MRS measurements of HTGC twice on the same day. Mean HTGC in the first $^1$H-MRS measurement was 4.0% and did not differ from the second $^1$H-MRS measurement, 4.0% (p=0.583). The CV was 4.5% and the RC was 0.4% HTGC (Table 1). The ICC between both scanning sessions was 0.999 (p<0.001).

Table 1: Summary of reproducibility statistics

<table>
<thead>
<tr>
<th></th>
<th>CV</th>
<th>RC</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) between weeks (n=24)</td>
<td>9.5%</td>
<td>1.3%</td>
<td>0.998</td>
</tr>
<tr>
<td>2) fatty liver (n=8)</td>
<td>4.1%</td>
<td>1.3%</td>
<td>0.997</td>
</tr>
<tr>
<td>3) within day (n=12)</td>
<td>4.5%</td>
<td>0.4%</td>
<td>0.999</td>
</tr>
</tbody>
</table>

CV = Coefficient of Variation  
RC = Repeatability Coefficient  
ICC = Intraclass Correlation Coefficient
DISCUSSION:

The reproducibility of $^1$H-MRS for measuring hepatic fat content was very acceptable for “between weeks” and “fatty liver” measurements. In this study “within liver” variability and “within day” reproducibility was acceptable. Assessment of “between weeks” reproducibility is essential to detect abnormal variation in longitudinal studies. There is sparse literature addressing “between weeks” reproducibility of $^1$H-MRS to measure HTGC. Longo et al. [22] studied reproducibility in two subjects on three consecutive days, with variability of 11% and 7%. From our data it can be concluded that normal variation in a single subject is lower than 1.3% HTGC. Interestingly the same variability is found for the subpopulation with fatty liver in this study, indicating the robustness of this technique. In addition, our data can be used to calculate sample sizes in intervention trials for detecting drug efficacy, but also for monitoring hepatic steatosis as a side effect of drug treatments.

We also studied reproducibility of two different voxel positions in the liver to investigate possible effect of heterogeneity of hepatic fat content in the liver on $^1$H-MRS measurements. Of interest, in our study the largest variation in HTGC % of $^1$H-MRS measurements was “within liver” variation. Our data are comparable to the existing literature. Johnson et al. [21] also found a CV of 14.5%, studied in five subjects. Szczepaniak et al. [1] found a lower CV of 11% studied 10 subjects. Thomas et al. [20] found a substantial 2 voxel inter individual variation (1 - 50%) in 12 volunteers. These results indicate a difference in hepatic fat content in different parts of the liver. In repeating $^1$H-MRS measurements of HTGC one should be aware of this and must perform measurements in the same voxel positions in the liver.

Furthermore, we found fairly good “within day” reproducibility, fitting within the range of CV’s reported in the literature. Johnson et al. [21] found a CV of 3.6% in five subjects, and Thomas et al. [20] a CV of 7% in 34 subjects. A higher CV was found by Szczepaniak et al. [1] in 10 subjects (CV of 8.5%). Machann et al. [27] studied reproducibility in five healthy subjects and found variation between 0.3% and 1.7%. Another study by Machann et al. [28] showed variations up to 10% in five healthy volunteers.

In our data, “within day” reproducibility is better than “between weeks” reproducibility. This finding suggests that variations are partly related to the MR scanner itself and partly to physiologic variation of liver fat. “Within day” reproducibility will be mainly related to measurement variability, whereas “between weeks” reproducibility contains both measurement variability and physiologic variations.
This study has some limitations. Although the number of patients was larger than in previous studies, still the number of patients is limited. As this was an exploratory pilot study no formal sample size calculation was carried out. We chose a sample size that was considered appropriate to address the study aims.

We did not compare our $^1$H-MRS results with reference standard liver biopsy due to its invasiveness. This would have been unethical in the majority of individuals in this study. Furthermore, we studied different aspects of reproducibility of $^1$H-MRS in different groups. Not all groups underwent all reproducibility investigations. However, our primary study aim - reproducibility of $^1$H-MRS for “between weeks” measurements - was performed in all subjects.

Standardization and lifestyle control was not implemented in this study. We think this is not strictly necessary since standardization and lifestyle control can be implemented in many different ways. In daily clinical practice patients are not standardized as well, and no lifestyle control is performed. Moreover, in clinical practice patients do not always obey rules on standardization and lifestyle control.

In this study, $^1$H-MRS is performed during free breathing. This is a potential limitation because the volume interrogated by $^1$H-MRS is blurred in the longitudinal direction by 2 to 3 centimetres respiratory excursions of the liver. In this study we did not encounter $^1$H-MRS acquisition problems caused by respiratory excursions. Voxels were carefully positioned in the right liver, avoiding the diaphragm by at least 4 cm. This way $^1$H-MRS was always performed in liver tissue. $^1$H-MRS during free breathing was not a disadvantage in this study.

Finally, it should be noted that in this study a 3.0 Tesla magnet of one vendor is used. It is unclear if these results can be extrapolated to 1.5 Tesla scanners and other vendors. No direct benefit of the increased spectral resolution at 3.0 Tesla is to be expected in HTGC measurements. It might even be that differences in e.g. shimming, B1-homogeneity, amount of eddy currents at 3.0 Tesla result in a less reproducible measurement (29). The same effects might play a role when comparing different scanners in different institutes and of different vendors. However, we do not expect these differences to be large, since the “within day” and “within liver” variability found in this study compares well with results obtained at 1.5 Tesla.

In conclusion, 3.0 Tesla $^1$H-MRS for the measurement of hepatic fat content is highly reproducible in a spectrum varying from low to high hepatic fat content. “Between weeks” reproducibility is clinically most relevant and displays a variation of 9.5%. We also showed that $^1$H-MRS is highly reproducible in subjects with fatty liver.
Because the variation in HTGC measurements between two voxel positions in the liver exceeds the "between weeks" variability, the same voxel position should be used in consecutive $^1$H-MRS measurements. Furthermore, "within liver" and "within day" reproducibility of 3.0 Tesla $^1$H-MRS to measure HTGC is comparable with reproducibility of $^1$H-MRS reported in the literature for 1.5 Tesla.

**Acknowledgments:**
We kindly thank Nikki Bodegom$^1$ for her contribution in performing the $^1$H-MRS scans.
$^1$BCN Neuroimaging Center, University Medical Center Groningen, Groningen, the Netherlands.
REFERENCES:


