R2* Relaxometry for the Quantification of Hepatic Iron Overload: Biopsy-Based Calibration and Comparison with the Literature

Abstract

Purpose: We compared the calibration of hepatic iron based on R2* relaxometry and liver biopsy with similar studies that have already been published to investigate the transferability of published calibration curves.

Materials and Methods: 17 patients with clinically suspected hepatic iron overload (HIO) were enrolled. All patients underwent liver biopsy and MRI of the liver using a multi-echo gradient echo sequence (TR = 200 ms; TE-initial 0.99 ms; Delta-TE 1.41 ms; 12 echos; flip-angle: 20°). R2* parameter maps were analyzed using manually placed regions of interest and R2* values were correlated with liver iron concentration (LIC) obtained from liver biopsy. In addition, the results of our study were compared with 6 similar, already published studies.

Results: A linear relationship between R2* and LIC was found. Regression analysis yielded a correlation coefficient of 0.926, a slope of 0.024 (mg/g) [95% CI 0.013–0.024] and an intercept of 0.277 (mg/g) [95% CI 0.328–2.49]. We found a significant correlation between the calibration curves obtained from our study in comparison to 3/6 similar studies. The other 3 studies used a different reference standard or sequence parameters which lead to a significant difference for slope, intercept or both in comparison to our data.

Conclusion: Calibration curves from published studies that are based on a correlation of liver biopsy and R2* can be used for the estimation of liver iron concentration, although different scanning parameters and post-processing protocols were used. Low initial TEs might be a prerequisite for pooling data for liver iron quantification.

Key Points:
- Calibration curves from different studies can be used for liver iron quantification
- For that purpose calibration curves from published studies should be based on liver biopsy
Introduction

Noninvasive assessment of hepatic iron overload is of increasing relevance because new oral iron chelators are available as an efficient treatment of systemic iron overload in patients with iron loading anemias and hepatic iron is a known risk factor for disease progression in different chronic liver diseases [1]. Hepatic iron overload (HIO) is a common complication in patients with thalassemia, sickle cell disease, aplastic anemia, myelodysplasia, hereditary hemochromatosis (HH) and non-alcoholic fatty liver disease [2]. Although elevated serum ferritin concentrations are frequently present in patients with iron overload, ferritin has a low sensitivity and specificity for hepatic iron overload [3, 4]. Liver iron concentration (LIC) closely correlates with the total iron stores, because the liver is the dominant iron storage organ [5, 6]. The reference method for the quantification of hepatic iron is chemical quantification of iron concentration in liver biopsies, but the limiting factors of this gold standard are that it is an invasive procedure with a 0.5 % hemorrhage risk [7]. Liver biopsies are also prone to sampling errors due to the inhomogeneous distribution of liver iron with 15 – 40 % coefficient of variation [8]. Magnetic resonance imaging (MRI) offers a noninvasive method to determine liver iron concentrations for the diagnosis of HIO or the noninvasive monitoring of phlebotomy and chelation therapy [2]. It is widely accepted for measuring liver and heart iron content in thalassemia patients, while data on the accuracy of MRI in other iron overload diseases is rather limited. Iron stored within the liver affects the magnetic resonance signal by altering the local magnetic field leading to a reduction of transverse relaxation times T2 and T2* [9]. Although this basic concept is simple, the technique is still, in many cases, managed by centers with expert radiologists and physicists. Different methods for magnetic resonance-based hepatic iron quantification are available but the major drawbacks are the significant cost of commercial protocols, scan time, post-processing with specialized software and the need for phantom studies. Numerous MRI methods have been proposed for liver iron evaluation and quantification [10]. Hepatic iron can be assessed by measuring signal intensity ratios (SIRs) of the liver and of a reference tissue with good correlation [11]. Direct measurement of the relaxation time (relaxometry) can also quantify liver iron by either measuring T2 relaxation times from spin echo sequences or T2* relaxation times (or relaxation rates R2*) from gradient echo sequences [4, 12]. Many studies have evaluated MRI for the quantification of hepatic iron by calibrating the used methods either against liver biopsy data, phantom data or even different MRI methods. [4, 11, 13 – 17]. At the same time many different acquisition protocols with variable scanning parameters and R2* estimation methods (magnitude vs. complex) have been used [18]. So far it is not clear if calibrations obtained in one study can easily be applied to other sites. Demonstration of the transferability of calibration data for noninvasive iron quantification protocols is required before widespread use of a specific protocol can be generally recommended. At our institute a multiecho T2* gradient echo (GRE) sequence is used to evaluate hepatic iron overload. In order to provide calibration for the hepatic iron estimation with our method, it was the purpose of this study to investigate the correlation between the obtained R2* values and the direct determination of hepatic iron concentration in liver biopsies. Furthermore to answer the question if published calibration curves could be used to implement hepatic iron estimation without direct validation by biopsy, we aimed to compare our results with results from similar studies that have already been published.

Materials and Methods

Patients

A total of 17 patients (3 women and 14 men; mean age 55, range 36 – 75) were enrolled between March 2003 and April 2014 at the Department of Radiology at the Medical University of Innsbruck. The patients were referred to MRI for iron quantification and underwent subsequent liver biopsy. In this retrospective study we identified patients who met the following inclusion criteria: (a) MRI of the liver accomplished with the sequence listed below, (b) liver biopsy with quantification of hepatic iron by open furnace atomic absorption spectrometry in air dried andashed biopsy samples, (c) time interval between MRI and biopsy < 60 days, (d) no therapy between liver biopsy and MRI examination, (e) no history of blood transfusion, (f) increased serum ferritin (> 300 μg/L in male patients and > 200 μg/L in female patients) or transferrin saturation (> 45 % in male patients and > 50 % in female patients). Written informed consent was obtained from each patient before performing MRI and liver biopsy. All patients were tested for the C282Y and H63D polymorphisms of the human hemochromatosis protein also known as the HFE gene. As a retrospective study, institutional review board approval was granted by means of a general waiver (local research ethics committee, Medical University of Innsbruck; 20 February 2009).

MRI protocol and post-processing

Magnetic resonance imaging was performed using a 1.5 T MR scanner (Magnetom Avanto, Siemens Healthcare Sector, Erlangen, Germany) using a body phased-array surface coil. R2* values were obtained using a fat-saturated (frequency selective fat saturation as provided by the manufacturer) multi-gradient echo sequence with 12 echoes (TR = 200 ms; TE-initial 0.99 ms; Delta-
The three ROI measurements. A summary of sequence details of all included studies.

Henninger B et al. R2* Relaxometry for... Fortschr Röntgenstr 2015; 187: 472–479

Results

The mean hepatic iron concentration was 4.947 mg/g, where absolute concentrations ranged from 0.917 mg/g to 11.646 mg/g (dry weight) in the entire patient cohort. MRI measurements of R2* ranged from 56.4 1/s to 471.6 1/s, and the mean was 191.8 1/s. 5 patients had HFE-associated hemochromatosis, 5 were classified as non-HFE hemochromatosis, 2 had a dysmetabolic iron overload syndrome (DIOS), 2 had aceruloplasminemia, 1 had spur cell anemia and 1 had sideroblastic anemia. The results are summarized in Table 2.

For the patients investigated at our department, we found a linear relationship between R2* measurements and LIC (Fig. 1). Regression analysis yielded a correlation coefficient of 0.926 (p < 0.001), a slope of 0.024 (mg/g) and an intercept of 0.277 (mg/g). The slope was significantly different from zero (p < 0.0001), whereas no significant difference from zero was found for the intercept (p = 0.645).

We were able to digitize 27/27 patients from Fig. 1 of the Anderson study, 42/43 from Fig. 1a of Hankins, 27/27 from Fig. 3 of Virtanen, 22/22 from Fig. 1 of Wood, 64/94 from Fig. 1 of Christoforidis and 50/50 from Fig. 2a of Garbowski [12 – 15, 17, 21]. The 30/94 patients from Fig. 1 of the study by Christoforidis were not digitized as they were declared in the paper to correspond to MR-HIC values equal to an upper limit of 250 μmol/g (shown as triangles in Fig. 1 of Christoforidis). The data point of one patient in Fig. 1a of Hankins coincided with the data point of another patient and could therefore not be separated.

Fig. 2 shows the pairwise comparison of our study with results from the studies mentioned above. Results of the linear regression analysis of all studies are summarized in Table 3. All datasets showed significant linear correlation. Comparing original and digitized data of the studies from Wood and Garbowski, we found an overall average deviation of only 0.358% (SD: 1.187%) for the digitized LIC values and 0.114% (SD: 0.868%) for the digitized R2* values. Also the comparison of the obtained fit curves for digitized and original data did not show any significant difference between slope and intercept (p: 0.966/0.994 for Wood and 0.933/0.981 for Garbowski).

There was no significant difference in slope and intercept between our data and data from Hankins, Wood and Garbowski. A significant difference for slope and intercept was found between our data and the data of Virtanen. The intercept of our data was significantly different from the study by Christoforidis and Anderson. The differences between our data and the published data are shown in Table 4.

Pooled data of 3 studies from the literature (Hankins, Wood and Garbowski) and the results of our study are shown in Fig. 3.

Discussion

In our study we found R2* values in the range of 56.4 1/s to 471.6 1/s. Linear regression resulted in an excellent correlation between R2* relaxation rate and results from liver biopsy for our setting with a correlation coefficient of 0.926. Thereby the linear relationship between R2* and the total LIC is consistent with ob-
Table 1: Summary of all examined studies.

<table>
<thead>
<tr>
<th>Study</th>
<th>Disease</th>
<th>Patient population and type of disease</th>
<th>R2* (ms)</th>
<th>TE initial (ms)</th>
<th>TE delta-TE (ms)</th>
<th>Number of echoes</th>
<th>Fitting method</th>
<th>Fat suppression</th>
<th>Correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Garbowski et al.</td>
<td>thalassemia major (n = 20)</td>
<td>multi-echo gradient echo; flip angle 20°; FOV 35 cm; matrix 96x128; 200 0.99 1.41 12 truncation exponential fit yes 0.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virtanen et al.</td>
<td>healthy patients (n = 6)</td>
<td>multi-echo gradient echo; flip angle 20°; FOV 35 cm; matrix 96x128; 200 0.99 1.41 12 truncation exponential fit yes 0.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hankins et al.</td>
<td>sickle cell disease (n = 10)</td>
<td>multi-echo gradient echo; flip angle 20°; FOV 48x48; 83 kHz bandwidth; SL 15 mm; matrix 64x64; 25 0.8 0.25 16 magnitude; exponential + c no 0.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Christoforidis et al.</td>
<td>thalassemia major (n = 32)</td>
<td>multi-echo gradient echo; flip angle 20°; FOV 35 cm; matrix 96x128; 200 0.99 1.41 12 truncation exponential fit yes 0.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Henninger et al.</td>
<td>(our study)</td>
<td>multi-echo gradient echo; flip angle 20°; FOV 35 cm; matrix 96x128; 200 0.99 1.41 12 truncation exponential fit yes 0.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Summary of all examined studies.

Table 2

<table>
<thead>
<tr>
<th>patient population and type of disease</th>
<th>R2* (ms)</th>
<th>TE initial (ms)</th>
<th>TE delta-TE (ms)</th>
<th>Number of echoes</th>
<th>Fitting method</th>
<th>Fat suppression</th>
<th>Correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wood et al.</td>
<td>200 2.2</td>
<td>2.24</td>
<td>8</td>
<td>magnitude; exponential + c no 0.68 (fibrotic samples)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hankins et al.</td>
<td>0.8 0.25</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virtanen et al.</td>
<td>120 4</td>
<td>2.24</td>
<td>8</td>
<td>non-linear regression fit only in-phase echoes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garbowski et al.</td>
<td>200 0.99</td>
<td>1.41</td>
<td>12</td>
<td>truncation exponential fit yes 0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.
the effects of noise on the T2* signal decay and evaluated different curve fitting models [19]. They concluded that the truncation model, which is also used in our study, proves to be reproducible and more accurate than the other used models (monoexponential, baseline subtraction and offset). The study by Meloni et al. also compared different post-processing approaches in R2* measurement, a single-exponential model fit and an exponential-plus-constant model fit [20]. They found large systematic differences at higher R2* values with the exponential-plus-constant fits averaging ~20 % higher. By using technique-appropriate calibration curves, this bias effectively disappeared, producing excellent agreement between the two approaches, so that it can be concluded that both signal decay models yield clinically acceptable estimates of LIC. Furthermore, it is known that differences in biopsy handling (paraffin-embedded versus fresh specimens) can influence the LIC and therefore the comparability of calibration curves [24].

When comparing our results with the study by Anderson et al., we found a significant difference compared to our study concerning intercept. This difference may be caused by systematic differences in liver biopsy, MRI acquisition (too long first TE) and post-processing. Probably small differences between studies can be attributed to sampling errors due to heterogeneous liver iron deposition [25]. Another explanation may be found in the different scanner types that were used. Westwood et al. found slightly higher T2* relaxation times of the heart in normal subjects with a scanner from the same manufacturer as the one used in the An-
Fig. 1  Relationship between $R_2^*$ (1/s) and liver iron concentration (mg/g) for the patients investigated at our department. The solid line represents the fitted linear regression model (slope = 0.024 s mg/g, intercept = 0.277 mg/g). The dotted lines represent the 95% confidence interval of the linear regression.

Fig. 2  Pairwise comparison of our study with results from other studies. The solid lines represent the fitted linear regression model for our data, with the dotted lines showing the corresponding 95% confidence interval. The dashed lines correspond to the linear regression model for the respective study. Maximum values have been adapted for direct visual comparison.
derson study compared to a scanner that was used in our study [26]. Virtanen et al. compared R2* measurements with iron estimates based on the SIR method and obtained a highly linear correlation (r = 0.981) [15]. Comparing our data with this study, however, we also found a significant difference between the calibration curve given by Virtanen et al. and our data. In accordance with other studies, our data indicate that the SIR method apparently overestimates LIC [23]. This may be crucial when therapeutic decisions are based on the results from noninvasive hepatic iron quantification. Christoforidis et al. compared the data of three different MR protocols (R2* relaxometry with gradient-echo sequences, SIR method and R2) in the assessment of liver iron content [17]. A good correlation between liver R2* and R2 measurement (r = 0.886) and between the R2 and the SIR method (r = 0.927) was found. However, also in this study no validation by biopsy was performed and we again found a significant difference between the given calibration curve and our data. Based on the studies of Christoforidis et al. and Virtanen et al., where liver iron content was estimated mainly by the SIR method, it appears that there is a significant difference in calibration curves when hepatic iron content was not determined by biopsy. Furthermore, all 3 studies (Anderson, Virtanen, Christoforidis) were different from our study regarding the initial TE. Our study and also the study by Wood, Hankins and Garbowski had an initial TE around 1 (0.93 – 1.1 ms). The other 3 studies used an initial TE of 2.2 ms and 4 ms which might also contribute to the observed significant difference of the calibration curves. Tanner et al. validated the transferability of the T2* technique for the quantification of tissue iron in a multi-center study [27]. The inter-center reproducibility of T2* in the heart and liver was 5.0% and 7.1%, with mean absolute differences of 1.3 ms and 0.45 ms. They concluded that the T2* technique is transferable between MR scanners with good reproducibility. Inter-study reproducibility has also been evaluated by Westwood et al., but only for T2* sequences for the heart – high transferability between scanners from different manufactures and between different sites could be demonstrated [26]. Our present data seem to support this finding. We found good agreement between calibration curves of different studies in the case where a validation by biopsy was performed suggesting that in this case calibration data might be pooled, as done in Fig. 3.

A limitation of our study is the small patient population. At our institution liver biopsy has strict indications and MRI is accepted as a tool for liver iron estimation. Therefore, it is not considered ethical to perform biopsy in every patient suspected to have liver iron overload especially when HFE testing is negative. Another limitation of our method could be that it is an ROI-based method. Meloni et al. evaluated the effectiveness of the single ROI approach and found that it slightly underestimates liver iron quantification due to susceptibility artifacts when the ROI was placed over segments VII and VIII [28]. They concluded that the single ROI approach can be safely used in the clinical area when taking care to avoid susceptibility artifact effect. McCarville et al. evaluated 41 patients with iron overload by R2* mapping and correlated the results with liver biopsy [16]. In their study they compared small ROI placement and a whole liver method. They found a strong correlation between liver iron content and R2* measurement by small ROI and whole liver ROI method, although they found slightly greater inter-observer variability when using the small ROI technique. In our study ROIs were placed in 3 positions: one in the left liver lobe and two in the right lobe. Due to image acquisition in 5 different transverse sections, we could place optimal ROIs to avoid possible artifacts due to breathing and vessels. Other study groups also used a global whole liver method with promising measurements and lower inter-observer variability [29]. Currently there is no full consensus on the best approach for the type of measurement. As another limitation, it should be noted that for comparison with other published studies we were able to obtain original data for only two of the studies and that the data had to be manually digitized for the remaining studies. However, as we found only a very small deviation between digitized and original data (below 1%) and no significant differences in slope and intercept for the respective fit curves, it seems that the use of digitized data for comparison with published studies is justifiable.

Finally in our study we used a multi-gradient echo sequence with spectral fat saturation which might influence the R2* values in the presence of high iron concentration due to e.g. line broadening of the water signal. Without spectral fat saturation a complex fitting...
model would have to be used taking multiple fat peaks into account which was beyond the scope of the present article. It should, however, be noted that we observed a very close relationship between our data and the data of Hankins (see Fig. 2), where no fat saturation with otherwise similar parameters was used. This might be an indication that the use of fat saturation did not introduce a clinically significant bias to our data. It will be the aim of our planned future studies to evaluate these topics further.

**Conclusion**

In conclusion, MRI at 1.5 T is ubiquitous and R2* relaxometry can be implemented for many scanners. Our study shows that calibration curves from published studies that are based on liver biopsy can be used for the estimation of liver iron concentration, although different scanning parameters and post-processing protocols were used. The 3 studies with the best agreement used initial TEs of around 1 ms. Low initial TEs might be a prerequisite for pooling data for liver iron quantification. With this, R2* can provide liver iron estimates throughout the clinically relevant range and could make hepatic iron quantification widely accessible without the need of individual biopsy-based calibration. Nevertheless, more studies especially with a focus on multi-center studies should be performed to study the topic of transferability of MRI-based hepatic iron estimation.

**Clinical Relevance of the Study**

- Calibration curves based on liver biopsy with R2* relaxometry from different studies can be transferred.
- The recommended initial TE of the used GRE sequence should be around 1 ms or below.
- Direct validation by own biopsy results is not mandatory.

**Acknowledgments**

The authors thank John C. Wood (Division of Cardiology, Children's Hospital of Los Angeles, CA) and Maciej W. Garbowski (University College London, UK) for providing their original data (R2* and LIC values).

**References**