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

R2*-Relaxometrie für die Quantifizierung einer Eisenüberladung der Leber: Biopsie-basierte Kalibrierung und Vergleich mit der Literatur

Authors   B. Henninger[1], H. Zoller[2], S. Rauch[1], A. Finkenstedt[2], M. Schocke[1], W. Jaschke[1], C. Kremser[1]

Affiliations   [1] Radiology, Medical University of Innsbruck, Austria
[2] Internal Medicine, Medical University of Innsbruck, Austria

Correspondence
Dr. Benjamin Henninger
Radiologie, Medizinische
Universität Innsbruck
Anichstraße 35
6020 Innsbruck
Austria
Tel.: ++43/5 12/50 48 09 14
Fax: ++43/5 12/50 42 54 98
benjamin.henninger@i-med.
ac.at

Zusammenfassung
▼

Ziel: Ziel dieser Studie war es, die Kalibrierung des Lebereisens basierend auf R2*-Relaxometrie und Leberbiopsie mit ähnlichen, bereits veröffentlichten, Studien zu vergleichen um die Übertragbarkeit der veröffentlichten Eichkurven zu untersuchen.

Material und Methoden: 17 Patienten mit klinischem Verdacht auf pathologischer Eiseneinlagerung wurden in diese Studie eingeschlossen. Bei allen Patienten wurden eine Leberbiopsie und eine MRT der Leber zur Quantifizierung des Gewebeeisens durchgeführt. Die R2*-Relaxationszeit wurde in identischen ROIs mit einer Fett-gesättigten Multi-Echo-Gradienten-Echo-Sequenz mit 12 Echos gemessen (TR = 200 ms; TE-initial 0,99 ms; Delta-TE 1,41 ms; 12 echos; flip-angle: 20 °). Diese Ergebnisse wurden den aus der Biopsie gewonnenen Eisenkonzentrationen gegenübergestellt. Als Weiteres wurden die Ergebnisse unserer Studie mit 6 ähnlichen Studien verglichen.

Ergebnisse: Es konnte ein linearer Zusammenhang zwischen R2* und der Lebereisenkonzentration gezeigt werden. Die Regressionsanalyse ergab einen Korrelationskoeffizienten von 0,926, eine slope (Steigung) von 0,024 (s mg/g) [95 % CI 0,013 – 0,024] und einen intercept von 0,277 (mg/g) [95 % CI –0,328 – 2,49]. Eine signifikante Korrelation zwischen der Eichkurve unserer Studie und jenen bei 3/6 ähnlichen Studien konnte gezeigt werden. Die anderen 3 Studien verwendeten einen unterschiedlichen Referenzstandard oder Sequenzparameter, die im Vergleich mit unserer Studie zu einem signifikanten Unterschied in Steigung, Achsenabschnitt oder beiden führten.

Schlussfolgerung: Kalibrierungskurven von veröffentlichen Studien, die auf einer Korrelation von Leberbiopsie und R2* basieren, können für die Schätzung der Lebereisenkonzentration verwendet werden, obwohl verschiedene Scanparameter und Nachbearbeitungsprotokolle verwendet wur-

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 (s 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

den. Niedrige initiale TEs könnten eine Voraussetzung für eine Zusammenlegung von Daten für die Quantifizierung von Lebereisen sein.

Kernaussagen:
▶ Eichkurven aus verschiedenen Studien können für die Quantifizierung von Lebereisen verwendet werden
▶ Zu diesem Zweck sollten Eichkurven aus veröffentlichten Studien auf einer Leberbiopsie basieren
▶ Niedrige initiale TEs könnten eine Voraussetzung für eine Zusammenlegung von Daten für die Quantifizierung von Lebereisen sein

▶ Low initial TEs might be a prerequisite for pooling data for liver iron quantification

Citation Format:
▶ Henninger B., Zoller H., Rauch S. et al. R2* Relaxometry for the Quantification of Hepatic Iron Overload: Biopsy-Based Calibration and Comparison with the Literature. Fortschr Röntgenstr 2015; 187: 472–479

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 and ashed 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 H63 D 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-

TE 1.41 ms; 12 echoes; flip-angle: 20°). During one breath-hold a single slice with a 10-mm slice thickness was acquired in transverse orientation and the acquisition was repeated for five different slice positions. The matrix was held constant at 128 × 128 pixels with a field of view of 380 × 380 mm.

Image analysis was performed independently by a radiologist (ROI placement) and a physicist (calculation of R2* maps). Offline post-processing included quantitative image analysis using ImageJ (Wayne Rasband, National Institutes of Health, Bethesda, MD, USA). R2* maps were calculated from the magnitude images by pixel-wise fitting with a truncation model [19] using a custom-written ImageJ plugin. Later echo times were manually excluded from the fit when the signal in the respective image dropped below the noise level and stayed approximately constant for further echo times [20]. Three regions of interest (ROI) were placed in the liver parenchyma of one transverse section (two in the right lobe and one in the left lobe). ROIs had a diameter of 10 to 13 mm (area between 0.8 and 1.3 cm^2) and were carefully placed to avoid major vessels. The mean R2* was calculated using the three ROI measurements.

Liver biopsy

Liver biopsy was performed only in patients who had a clinical indication. Ultrasound-guided percutaneous liver biopsy was carried out using a 16-gauge true-cut biopsy needle. One complete core was sent for iron quantification natively in a trace element-free container. For liver iron quantitation the samples were sent according to our standard clinical practice to an external, certified laboratory where iron was quantified with graphite furnace atomic absorption spectrometry. The LIC (mg Fe/g liver dry weight) was reviewed by one pathologist blinded to MRI results.

Comparison between R2* times, biopsy results and the literature

Results of the R2* measurements (1/s) were correlated with results from liver biopsy (mg/g). To compare our results with data from the literature, to the best of our knowledge, only 6 studies in which calibration curves (correlation between R2* and LIC) are shown (Wood et al., Anderson et al., Hankins et al., Virtanen et al., Christoforidis et al. and Garbowski et al.) could be identified [12 – 15, 17, 21]. For two of these studies the original data were generously made available to us (Wood, Garbowski). For the remaining four studies the published data were digitized using ImageJ. Thereby a digital image (screen capture) of the respective calibration curve was loaded into the software and the x, y coordinates of the individual data points were listed after manual placement of the point selection tool. Taking into account the coordinates and corresponding values of the individual corner points of the diagram, the values of each data point were calculated. To obtain an indicator for the accuracy of this digitalization procedure, this was also performed for the studies of Wood and Garbowski where a direct comparison with the original data was possible. In two studies (Virtanen et al., Christoforidis et al.) the LIC was given in units of µmol/g and had to be converted to units of mg/g using: Fe [mg/g] = 55.845*10^{-3} * Fe [µmol/g]. ● **Table 1** summarizes sequence details of all included studies.

Statistical analysis

All statistical calculations were performed using the R Project for Statistical Computing [R Development Core Team (2006), Vienna, Austria, Version 2.13.1]. For linear regression analysis, a linear model was fitted to the data. To compare the obtained regression lines of our data with published data, analysis of covariance (ANCOVA) was used. The results were considered significant when the P-value was less than 0.05.

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 (s 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.

Tab. 1 Zusammenfassung aller untersuchten Studien.

study	patient population and type of disease	sequence (1.5 T) used to measure R2*	TR (ms)	TE-initial (ms)	delta-TE (ms)	number of echoes	fitting method	fat suppression	correlation coefficient
Anderson et al.	beta-thalassemia (n = 27)	multi-echo gradient echo; flip angle 20°; SL 10mm; matrix 96 × 128, FOV 35cm; sampling bandwith 125 kHz	200	2.2	2.24	8	magnitude; exponential + c	no	0.68 (fibrotic samples) 0.93 (non-fibrotic samples)
Wood et al.	22/102 patients had MRI and concomitant biopsy: thalassemia major (n = 9), sickle cell disease (n = 10), thalassemia intermedia (n = 2), Blackfan-Diamond syndrome (n = 1)	single-echo gradient echo; flip angle 20°; SL 15mm; matrix 64x64; FOV 48x24; 83 kHz bandwidth	25	0.8	0.25	16	magnitude; exponential + c	no	0.97
Hankins et al.	sickle cell anemia (n = 32), thalassemia major (n = 6), bone marrow failure (n = 5)	multi-echo gradient echo; SL 10mm	not mentioned	1.1	−17.3 (0.8)	20	truncated exponential fit	no	0.96 – 0.98
Virtanen et al.	healthy patients (n = 6), patients suspected for HIO (n = 27)	multi-echo gradient echo; flip angle 20°; SL 10mm; matrix 256x128; FOV 400 mm	120	4	irregular echo spacing (5.7)	4	non-linear regression fit	only in-phase echoes	0.98 (between R2* and SIR)
Christoforidis et al.	thalassemia patients (n = 94)	multi-echo gradient echo; flip angle 20°; FOV 35 cm	200	2.24	2.24	8	mono-exponential decay curve	not mentioned	0.85 (between R2* and SIR)
Garbowski et al.	thalassemia major (n = 20), Diamond-Blackfan anemia (n = 2), congenital sideroblastic anemia (n = 2), pyruvate kinase deficiency anemia (n = 1)	multi-echo single breath-hold gradient echo; SL 10 mm	not mentioned	0.93	0.8	20	truncation exponential fit	not mentioned	0.94
Henninger et al. (our study)	○ Table 2	multi-echo gradient echo; flip angle 20°;FOV 36x36; matrix 128x128	200	0.99	1.41	12	truncation exponential fit	yes	0.92

servations by other investigators [13]. For a widespread clinical application of the method, a linear relationship facilitates iron estimation based on measured R2* values. We found no significant difference between our calibration curve and the calibration curve obtained from the studies by Wood, Hankins and Garbowski. All have in common that the LIC was maintained by liver biopsy and that the scanning sequence had an initial TE of around 1 ms. There was a significant difference for slope, intercept or both between our calibration and the calibration curve of the 3 other studies (Virtanen, Christoforidis and Anderson) in which the initial TE was between 2.2 ms and 4 ms. In the study by Virtanen and Christoforidis the LIC was provided by other MRI techniques and not by liver biopsy.

St. Pierre et al. performed single spin-echo (SE) sequences and found a curvilinear relationship between R2 and HIC with a correlation coefficient of 0.98 [4]. This so-called "ferriscan" method is commercially available. The R2* method is a cost-effective alternative to SE imaging with "ferriscan" and can be carried out in a few minutes whereas R2 "ferriscan" takes longer to collect (20 minutes of acquisition). Juchems et al. compared the R2 method of St.Pierre with the SIR method [22]. The two methods result in different liver iron concentrations with generally higher values for the SIR method, which is based on GRE sequences. The correlation between both methods was still significant (r = 0.85). Castiella et al. recently demonstrated that the SIR method has a tendency to overestimate LIC and that it could measure iron up to 350 µmol/g (19.5 mg/g) [23]. None of our patients had an LIC at that level, although one patient with HFE-associated hemochromatosis had an LIC of 11.64 mg/g which is quite high for that disease.

The R2* (or T2*) method provides the possibility of transferability, no matter which scanner type is used [15]. In order to test this, we compared our results with 6 other studies. Wood et al. assessed 102 patients with 22 having concomitant liver biopsy [13]. They also found a close relationship between R2* and HIC (r = 0.97) and concluded that R2* can accurately measure hepatic iron with values in the entire range of iron overload. Hankins et al. investigated 44 patients with R2* and liver biopsy [14]. HIC and R2* MRI had a strong correlation with coefficients of 0.96 – 0.98. Garbowski et al. used an optimized T2* sequence calibrated against 50 liver biopsy samples on 25 patients with transfusional hemosiderosis [21]. They also found a near-linearity correlation between R2* and LIC (Pearson r = 0.94). There was no significant difference between the calibration curves obtained from our study and the data of the mentioned studies from Wood, Hankins and Garbowski, although R2* analysis methods and sequence parameters were slightly different with e. g. a different number of echoes. He at al. studied

Table 2   Patient population with MRI and biopsy data.

Tab. 2   Patientenpopulation mit MRT und Biopsiedaten.

no.	sex	age (years)	diagnosis	genetic testing	R2* (1/s)	LIC (mg/g)
1	m	57	dysmetabolic iron overload syndrome	–	56.4	0.92
2	m	50	non-HFE hemochromatosis	–	59.5	1.02
3	m	53	HFE-associated hemochromatosis	C282Y/H63 D compound heterozygosity	75.8	1.53
4	m	65	non-HFE hemochromatosis	–	96.8	1.42
5	m	51	HFE-associated hemochromatosis	C282Y/H63 D compound heterozygosity	101.1	2.38
6	m	74	non-HFE hemochromatosis	–	124.9	2.85
7	m	75	dysmetabolic iron overload syndrome	–	111.6	3.38
8	m	55	non-HFE hemochromatosis	–	161.1	3.50
9	m	44	non-HFE hemochromatosis	–	125.9	5.06
10	m	42	sideroblastic anemia	–	255.9	5.04
11	m	43	HFE-associated hemochromatosis	C282Y/H63 D compound heterozygosity	153.9	5.07
12	m	36	non-HFE hemochromatosis	–	213.6	5.57
13	f	68	spur cell anemia	–	143.2	5.99
14	f	43	aceruloplasminemia	–	321.6	7.81
15	m	67	HFE-associated hemochromatosis	C282Y homozygosity	471.6	9.84
16	f	41	aceruloplasminemia	–	437.4	11.12
17	m	71	HFE-associated hemochromatosis	C282Y homozygosity	350.1	11.65

Table 3   Results of linear regression analysis.

Tab. 3   Resultate der linearen Regressionsanalyse.

	slope (s mg/g)	95 % CI	std. error	p	intercept (mg/g)	95 % CI	std. error	p	r
Anderson (○ **Fig. 1a**)	0.017	0.012 – 0.022	0.002	< 0.0001	–0.347	–2.84 – 2.146	1.210	0.777	0.831
Wood (○ **Fig. 1**)	0.027	0.024 – 0.030	0.002	< 0.0001	–0.188	–2.7 – 2.323	1.204	0.877	0.970
Wood (original data)	0.027	0.024 – 0.031	0.002	< 0.0001	–0.259	–3.011 – 2.492	1.315	0.846	0.968
Hankins (○ **Fig. 1**)	0.027	0.025 – 0.03	0.001	< 0.0001	–0.294	–1.454 – 0.867	0.574	0.612	0.963
Virtanen (○ **Fig. 3**)	0.043	0.04 – 0.047	0.001	< 0.0001	–1.035	–1.625 – 0.445	0.286	< 0.002	0.981
Christoforidis (○ **Fig. 1**)	0.029	0.024 – 0.033	0.002	< 0.0001	3.286	1.721 – 4.852	0.783	< 0.0001	0.828
Garbowski (○ **Fig. 2a**)	0.032	0.027 – 0.037	0.002	< 0.0001	0.309	–2.126 – 2.743	1.218	0.8	0.896
Garbowski (orginal data)	0.032	0.028 – 0.037	0.002	< 0.0001	0.210	–2.235 – 2.656	1.216	0.863	0.896
our study	0.024	0.013 – 0.024	0.002	< 0.0001	0.277	–0.328 – 2.49	0.589	0.645	0.926

Table 4   Comparison between our data and different published data.

Tab. 4   Vergleich zwischen unseren Daten und verschiedenen veröffentlichten Daten.

	P (interaction/ difference between slopes)	P (difference of intercepts)
Anderson (○ **Fig. 1a**)	0.248	0.042
Wood (○ **Fig. 1**)	0.584	0.874
Wood (original data)	0.577	0.909
Hankins (○ **Fig. 1**)	0.386	0.873
Virtanen (○ **Fig. 3**)	< 0.0001	< 0.001
Christoforidis (○ **Fig. 1**)	0.53	< 0.0001
Garbowski (○ **Fig. 2a**)	0.344	0.222
Garbowski (original data)	0.36	0.21

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-
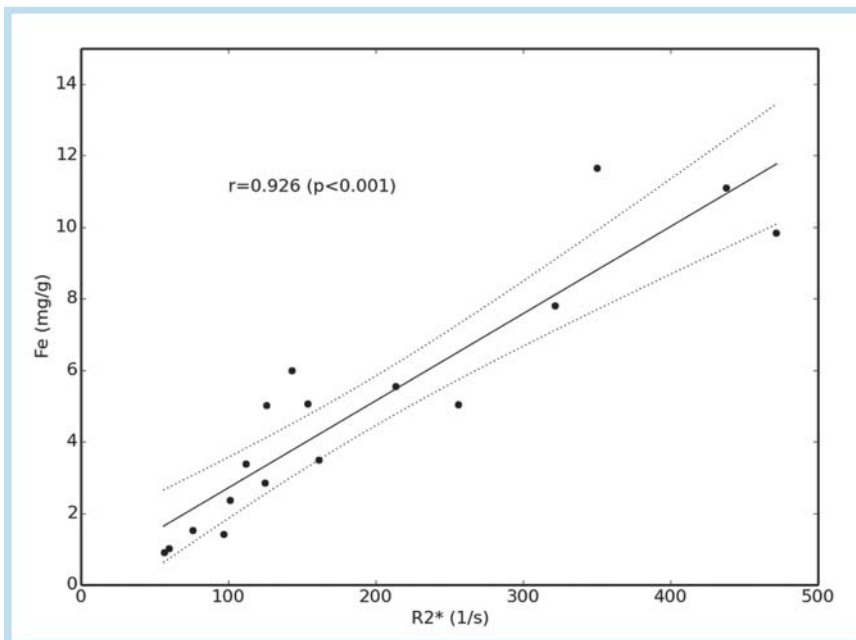
Fig. 1 Relationship between R2* (1/s) and liver iron concentration (mg/g) for the patients investigated at our department. The solid line represents the fitted linear regression model (slope = 0.024 s mg/g, intercept = 0.277 mg/g). The dotted lines represent the 95 % confidence interval of the linear regression.

Abb. 1 Zusammenhang zwischen R2* (1/s) und der Lebereisenkonzentration (mg/g) für die auf unserer Abteilung untersuchten Patienten. Die durchgezogene Linie stellt das angepasste lineare Regressionsmodell dar (Steigung = 0,024 s mg/g, Achsenschnittpunkt = 0,277 mg/g). Die gestrichelten Linien stellen den 95 % Konfidenzintervall der linearen Regression dar.
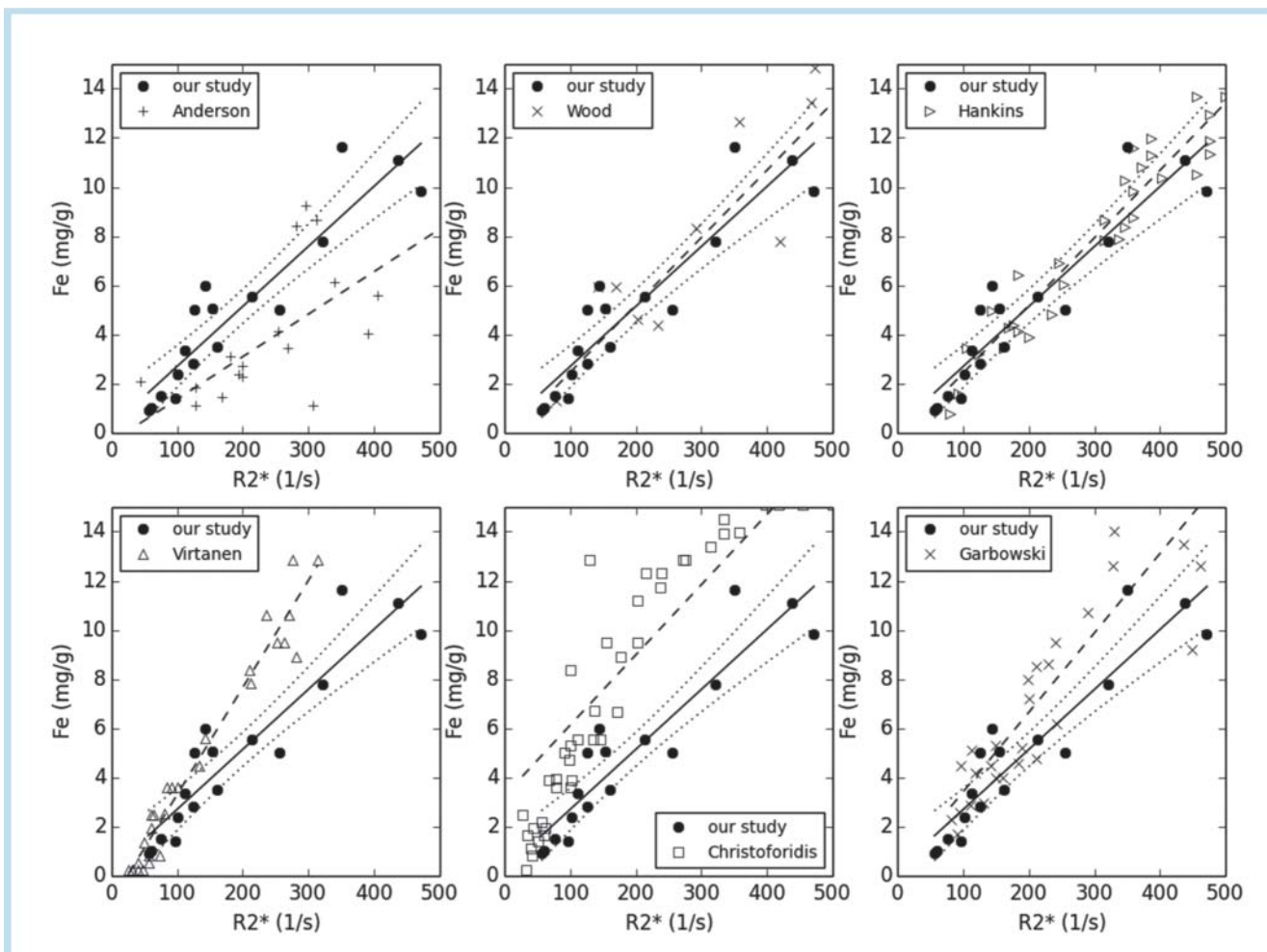

Fig. 2 Pairwise comparison of our study with results from other studies. The solid lines represent the fitted linear regression model for our data, with the dotted lines showing the corresponding 95 % confidence interval. The dashed lines correspond to the linear regression model for the respective study. Maximum values have been adapted for direct visual comparison.

Abb. 2 Paarweiser Vergleich unserer Studie mit den Ergebnissen anderer Studien. Die durchgezogenen Linien repräsentieren das angepasste lineare Regressionsmodell unserer Daten, die gestrichelten Linien entsprechen dem 95 % Konfidenzintervall. Die gestrichelten Linien entsprechen dem linearen Regressionsmodell für die jeweilige Studie. Die Höchstwerte wurden für einen direkten visuellen Vergleich angepasst.
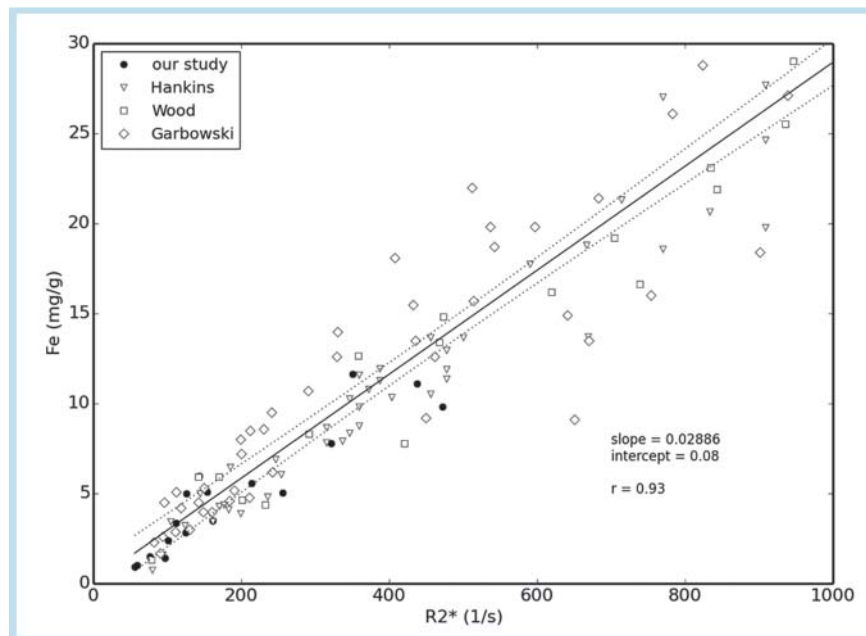
Fig. 3   Pooled data of 3 studies from the literature (13, 14, 21) and results of our study. The solid line represents the fitted linear regression model (slope = 0.02 876 s mg/g, intercept = 0.137 mg/g). The dotted lines represent the 95 % confidence interval of the linear regression.

Abb. 3   Zusammengefasste Daten der drei Studien aus der Literatur (13, 14, 21) und die Ergebnisse unserer Studie. Die durchgezogene Linie stellt das angepasste lineare Regressionsmodell dar (Steigung = 0,02 876 s mg/g, Achsenabschnitt = 0,137 mg/g). Die gestrichelten Linien stellen den 95 % Konfidenzintervall der linearen Regression dar.

derson study compared to a scanner from the manufacturer 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
▼

References

1 *Adams PC, Barton JC.* How I treat hemochromatosis. Blood 2010; 116: 317–325
2 *Sirlin CB, Reeder SB.* Magnetic resonance imaging quantification of liver iron. Magn Reson Imaging Clin N Am 2010; 18: 359–381, ix
3 *Brittenham GM, Cohen AR, McLaren CE et al.* Hepatic iron stores and plasma ferritin concentration in patients with sickle cell anemia and thalassemia major. Am J Hematol 1993; 42: 81–85
4 *St Pierre TG, Clark PR, Chua-anusorn W et al.* Noninvasive measurement and imaging of liver iron concentrations using proton magnetic resonance. Blood 2005; 105: 855–861
5 *Angelucci E, Brittenham GM, McLaren CE et al.* Hepatic iron concentration and total body iron stores in thalassemia major. N Engl J Med 2000; 343: 327–331
6 *Crisponi G, Ambu R, Cristiani F et al.* Does iron concentration in a liver needle biopsy accurately reflect hepatic iron burden in beta-thalassemia? Clin Chem 2000; 46: 1185–1188
7 *Janes CH, Lindor KD.* Outcome of patients hospitalized for complications after outpatient liver biopsy. Ann Intern Med 1993; 118: 96–98
8 *Villeneuve JP, Bilodeau M, Lepage R et al.* Variability in hepatic iron concentration measurement from needle-biopsy specimens. J Hepatol 1996; 25: 172–177
9 *Tang H, Jensen JH, Sammet CL et al.* MR characterization of hepatic storage iron in transfusional iron overload. J Magn Reson Imaging 2013
10 *Schonnagel BP, Fischer R, Nielsen P et al.* Iron quantification in iron overload disease using MRI. Fortschr Röntgenstr 2013; 185: 621–627
11 *Gandon Y, Olivie D, Guyader D et al.* Non-invasive assessment of hepatic iron stores by MRI. Lancet 2004; 363: 357–362
12 *Anderson LJ, Holden S, Davis B et al.* Cardiovascular T2-star (T2*) magnetic resonance for the early diagnosis of myocardial iron overload. Eur Heart J 2001; 22: 2171–2179
13 *Wood JC, Enriquez C, Ghugre N et al.* MRI R2 and R2* mapping accurately estimates hepatic iron concentration in transfusion-dependent thalassemia and sickle cell disease patients. Blood 2005; 106: 1460–1465
14 *Hankins JS, McCarville MB, Loeffler RB et al.* R2* magnetic resonance imaging of the liver in patients with iron overload. Blood 2009; 113: 4853–4855
15 *Virtanen JM, Komu ME, Parkkola RK.* Quantitative liver iron measurement by magnetic resonance imaging: in vitro and in vivo assessment of the liver to muscle signal intensity and the R2* methods. Magn Reson Imaging 2008; 26: 1175–1182
16 *McCarville MB, Hillenbrand CM, Loeffler RB et al.* Comparison of whole liver and small region-of-interest measurements of MRI liver R2* in children with iron overload. Pediatr Radiol 2010; 40: 1360–1367
17 *Christoforidis A, Perifanis V, Spanos G et al.* MRI assessment of liver iron content in thalassamic patients with three different protocols: comparisons and correlations. Eur J Haematol 2009; 82: 388–392
18 *Hernando D, Kramer JH, Reeder SB.* Multipeak fat-corrected complex R2* relaxometry: theory, optimization, and clinical validation. Magn Reson Med 2013; 70: 1319–1331
19 *He T, Gatehouse PD, Smith GC et al.* Myocardial T2* measurements in iron-overloaded thalassemia: An in vivo study to investigate optimal methods of quantification. Magn Reson Med 2008; 60: 1082–1089
20 *Meloni A, Rienhoff HY Jr, Jones A et al.* The use of appropriate calibration curves corrects for systematic differences in liver R2* values measured using different software packages. Br J Haematol 2013; 161: 888–891
21 *Garbowski MW, Carpenter JP, Smith G et al.* Biopsy-based calibration of T2* magnetic resonance for estimation of liver iron concentration and comparison with R2 Ferriscan. J Cardiovasc Magn Reson 2014; 16: 40
22 *Juchems MS, Cario H, Schmid M et al.* Liver iron content determined by MRI: spin-echo vs. gradient-echo. Fortschr Röntgenstr 2012; 184: 427–431
23 *Castiella A, Alustiza JM, Emparanza JI et al.* Liver iron concentration quantification by MRI: are recommended protocols accurate enough for clinical practice? Eur Radiol 2011; 21: 137–141
24 *Butensky E, Fischer R, Hudes M et al.* Variability in hepatic iron concentration in percutaneous needle biopsy specimens from patients with transfusional hemosiderosis. Am J Clin Pathol 2005; 123: 146–152
25 *Emond MJ, Bronner MP, Carlson TH et al.* Quantitative study of the variability of hepatic iron concentrations. Clin Chem 1999; 45: 340–346
26 *Westwood MA, Anderson LJ, Firmin DN et al.* Interscanner reproducibility of cardiovascular magnetic resonance T2* measurements of tissue iron in thalassemia. J Magn Reson Imaging 2003; 18: 616–620
27 *Tanner MA, He T, Westwood MA et al.* Multi-center validation of the transferability of the magnetic resonance T2* technique for the quantification of tissue iron. Haematologica 2006; 91: 1388–1391
28 *Meloni A, Luciani A, Positano V et al.* Single region of interest versus multislice T2* MRI approach for the quantification of hepatic iron overload. J Magn Reson Imaging 2011; 33: 348–355
29 *Positano V, Salani B, Pepe A et al.* Improved T2* assessment in liver iron overload by magnetic resonance imaging. Magn Reson Imaging 2009; 27: 188–197