


# Comparison of clinical MRI liver iron content measurements using signal intensity ratios, $R_2$ and $R_2^*$

Jurgen H. Runge <sup>1</sup>, Erik M. Akkerman,<sup>1</sup> Marian A. Troelstra,<sup>1</sup> Aart J. Nederveen,<sup>1</sup> Ulrich Beuers,<sup>2</sup> Jaap Stoker<sup>1</sup>

<sup>1</sup>Department of Radiology, Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105AZ Amsterdam, The Netherlands

<sup>2</sup>Department of Gastroenterology & Hepatology, Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105AZ Amsterdam, The Netherlands

## Abstract

**Purpose:** To compare three types of MRI liver iron content (LIC) measurement performed in daily clinical routine in a single center over a 6-year period.

**Methods:** Patients undergoing LIC MRI-scans (1.5T) at our center between January 1, 2008 and December 31, 2013 were retrospectively included. LIC was measured routinely with signal intensity ratio (SIR) and MR-relaxometry ( $R_2$  and  $R_2^*$ ) methods. Three observers placed regions-of-interest. The success rate was the number of correctly acquired scans over the total number of scans. Interobserver agreement was assessed with intraclass correlation coefficients (ICC) and Bland–Altman analysis, correlations between  $LIC_{SIR}$ ,  $R_2$ ,  $R_2^*$ , and serum values with Spearman's rank correlation coefficient. Diagnostic accuracies of  $LIC_{SIR}$ ,  $R_2$  and serum transferrin, transferrin-saturation, and ferritin compared to increased  $R_2^*$  ( $\geq 44$  Hz) as indicator of iron overload were assessed using ROC-analysis.

**Results:** LIC MRI-scans were performed in 114 subjects. SIR,  $R_2$ , and  $R_2^*$  data were successfully acquired in 102/114 (89%), 71/114 (62%), and 112/114 (98%) measurements, with the lowest success rate for  $R_2$ . The ICCs of SIR,  $R_2$ , and  $R_2^*$  did not differ at 0.998, 0.997, and 0.999.  $R_2$  and serum ferritin had the highest diagnostic accuracies to detect elevated  $R_2^*$  as mark of iron overload.

**Conclusions:** SIR and  $R_2^*$  are preferable over  $R_2$  in terms of success rates.  $R_2^*$ 's shorter acquisition time and wide range of measurable LIC values favor  $R_2^*$  over SIR for MRI-based LIC measurement.

**Key words:** Magnetic resonance imaging—Iron overload—Hemochromatosis—Blood transfusion—Biomarker—Relaxometry

## Abbreviation

LIC Liver iron content

Various diseases are associated with increased liver iron content (LIC), which may induce or contribute to liver damage [1–3]. Serial measurement of LIC during long-term follow-up and treatment is highly desirable, but repeated invasive measurements are not recommended due to risks of complications of serial liver biopsies. Surrogate biochemical markers including serum ferritin and transferrin-saturation are widely used, but are flawed by limited specificity. Thus, accurate non-invasive MRI-based methods of LIC measurement are used in clinical practice for patients (suspected) with increased LIC [4, 5].

Several types of MRI LIC measurement have been described in the literature. Straightforward in–out phase gradient echo (GRE) shows signal loss at the later echo time (TE) but is only qualitative and easily confounded by the presence of hepatic steatosis. Quantitative approaches include (i) signal intensity ratio (SIR) measurement (e.g., the Gandon method) and (ii) MR-

**Electronic supplementary material** The online version of this article (doi:10.1007/s00261-016-0831-7) contains supplementary material, which is available to authorized users.

Correspondence to: Jurgen H. Runge; email: j.h.runge@amc.uva.nl

relaxometry. The Gandon method (henceforth referred to as “SIR”) utilizes the liver-to-muscle SIR on differently weighted MRI-scans [6]. This method allows easy and free calculation of the  $LIC_{SIR}$ , by entering ROI values in an online tool [7]. Hence, assuming the acquisition and placement of regions-of-interest (ROIs) are performed correctly, the method is robust to observer influences. A major limitation is its upper limit of detection of 350  $\mu\text{mol/g}$  (equal to 20 mg/g): changes above that threshold cannot be measured.

MR-relaxometry relies on the calculation of tissue relaxation rates ( $R_2$  and  $R_2^*$ , the inverse of relaxation times  $T_2$  and  $T_2^*$ ), which increase as iron accumulates and are sensitive to changes in LIC values well above the SIR-threshold. One commercialized  $R_2$  approach using single-echo spin-echo (SE) MRI is the FDA-approved St. Pierre method [FerriScan<sup>®</sup>], performed in 10 min in free-breathing [8]. The per-scan analysis price is ~\$300, on top of the costs of the MRI-scan itself. Alternative free-of-charge approaches are available for  $R_2$  using free-breathing or respiratory triggered SE-MRI and for  $R_2^*$  using single breath-hold GRE MRI [9].

Recent developments in MR-relaxometry include multiplex fat corrections and the use of complex instead of magnitude-only data fitting [10], assessment of the effect of fat suppression on  $R_2^*$  [11] and the comparison of advanced data fit models [12] and analysis approaches [13].

A comparative study of  $LIC_{SIR}$ ,  $R_2$ , and  $R_2^*$  in 94 patients with  $\beta$ -thalassemia reported high correlations [14]. However, success rates, interobserver agreement, and applicability for diseases other than  $\beta$ -thalassemia were not investigated, nor were serum markers assessed. The latter may be useful to screen for elevated LIC (i.e., >36  $\mu\text{mol/g}$ ), saving expensive and limited MRI time. We hypothesize that  $R_2^*$  is preferable over SIR and  $R_2$  in terms of success rate, acquisition time, and range of detection and over serum values in terms of accuracy in detecting elevated LIC.

In our center, the clinical LIC protocol has included SIR,  $R_2$ , and  $R_2^*$  since 2005, with regular weekly clinical referrals since 2008. The SIR measurement is recommended by the national guideline for hemochromatosis [15]. It is supplemented by  $R_2$  and  $R_2^*$  measurements to fill the gap caused by the SIR method’s hard cut-off at 350  $\mu\text{mol/g}$ . To investigate our hypothesis, we (i) assessed SIR,  $R_2$ , and  $R_2^*$  LIC measurements and their success rates and interobserver agreement; and (ii) compared the diagnostic accuracies of  $LIC_{SIR}$ ,  $R_2$ , and surrogate serum markers for correctly predicting elevated LIC based on increased  $R_2^*$ .

## Materials and methods

### *Ethical*

All data used for this study were acquired in clinical setting and were anonymized prior to analysis. Informed

consent was waived by the Medical Research Ethics Committee of the AMC Amsterdam.

### *Patients*

All MRI-based LIC measurements performed between January 1, 2008 and December 31, 2013 were retrospectively included in this study. As additional measurements were added to the protocol in 2014, only measurements up to end 2013 were included. Clinical diagnosis and—when available—serum markers of iron metabolism (total iron, transferrin, transferrin-saturation, ferritin) were collected and subsequently anonymized by a colleague not otherwise involved in this study.

### *MRI*

MRI-scanning was performed supine, feet first on a 1.5T Avanto MRI-scanner (Siemens AG, Erlangen, Germany) using phased-array coils (body array and spine coil) for localizers and  $R_2$  and  $R_2^*$  measurements and the body coil for the SIR measurement [6]. Use of the body coil provided an as homogenous  $B_1$  field as possible, reducing variation in SIR measurements due to variations of flip angles between patients. For  $R_2^*$  and  $R_2$ , the  $B_1$  variation is eliminated via the data fit. Breath-hold imaging (localizers, SIR and  $R_2^*$ ) was performed in expiration. Three 10-mm slices with a variable slice gap to cover the liver were equally positioned for all three LIC measurements. Especially for the GRE-based SIR and  $R_2^*$  measurements, careful  $B_0$  shimming is important to achieve a homogenous  $B_0$  field, ensuring correct measurements. Shimming was performed with a shim box covering the field-of-view in the feet-head direction and the contours of the abdomen (i.e., excluding the arms) in the left–right and anterior-posterior directions. The SIR measurement according to Gandon et al. requires five (T1, PD, T2, T2+, and T2++) image weightings with specific TR/TE combinations [6]. Table 1 contains an overview of the relevant scan parameters. Of note, the TE interval used for  $R_2^*$  was shorter (1.41 ms) than the standard in- and out-of-phase interval (2.26 ms).

### *Data analyses*

After inclusion all measurements were checked for correct TRs, TEs, and RF coils using DICOM header information as for SIR measurements, specific TR/TE combinations and the use of the body coil are mandatory. Image quality was assessed by a research trainee (JHR, 4 years of experience) and an abdominal radiologist (JS, 20 years of experience) using a 3-point scale (good/adequate/inadequate). The type of artifact(s) was noted. Measurements with incorrect scan parameters or inadequate image quality were classified unsuccessful.

**Table 1.** MRI parameters

	SIR	$R_2$	$R_2^*$
Technique	GRE	SE	GRE
TR (ms)	120	3000–4000 <sup>a</sup>	300
TE1 (ms)	Variable [7]	6.2	0.99
$\Delta$ TE (ms)	n/a	6.2	1.41
Number of echoes	1	16 (multiecho)	12 (multiecho)
FA (°)	Variable [7]	180	20
FOV (mm × mm)	380 × 285	380 × 285	380 × 285
Acquisition matrix	256 × 256	256 × 256	128 × 96 <sup>b</sup>
Reconstruction matrix	256 × 192	256 × 192	256 × 192
Parallel imaging	No	GRAPPA	GRAPPA
Acceleration factor	n/a	2	2
Bandwidth (Hz/pixel)	140	465	1963
Slice thickness	10	10	10
Slice gap	Variable <sup>c</sup>	Variable <sup>c</sup>	Variable <sup>c</sup>
Number of slices	3	3	3
Acquisition time	100 s (5 × 20 s)	9–16 min <sup>a</sup>	20 s

<sup>a</sup> Depending on the patient's respiratory frequency: one TR per respiratory cycle

<sup>b</sup> Zero-padding was used to fit  $R_2^*$  acquisition in breath-hold time (20 s)

<sup>c</sup> The slice gap was adjusted per patient so as to cover the whole liver with the three slices

### ROI-placement

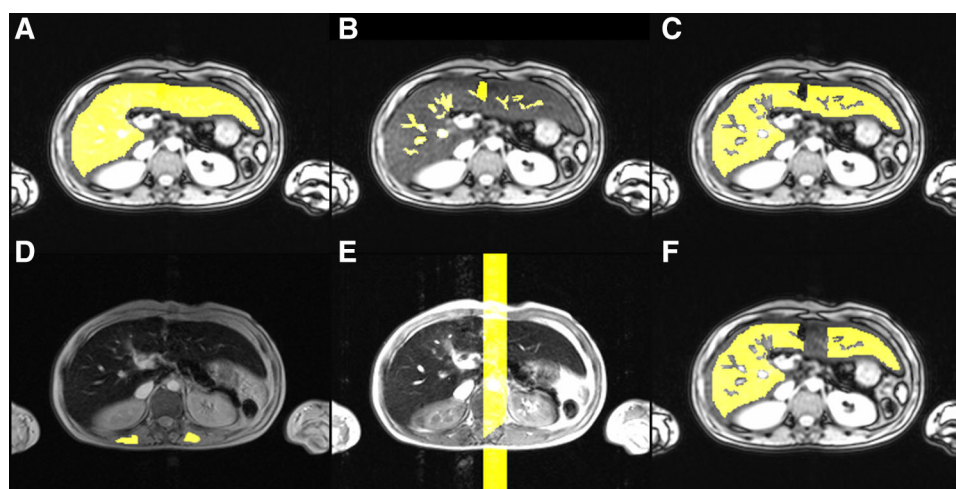
SIR,  $R_2$ , and  $R_2^*$  data were processed using custom-made software that allowed ROI-placement,  $LIC_{SIR}$  calculation, and  $R_2$  and  $R_2^*$  data fitting. Three blinded observers (JHR, MAT, and EMA) with four, a half and 9 years of experience, respectively, independently placed regions-of-interest (ROIs) for three slices per scan. First, the liver parenchyma was masked on  $R_2^*$  source data, excluding a rim near the liver edge (Fig. 1A). Next, non-liver voxels (e.g., vessels, gall bladder) inside the liver contour were masked (Fig. 1B). By subtracting ROI-2

from ROI-1, only liver parenchyma remained (Fig. 1C). Liver ROIs were copied from the  $R_2^*$  data for SIR analysis, with two additional ROIs in both paraspinal muscles, carefully avoiding areas of signal intensity loss close to the lung (Fig. 1D). This also allowed a check to identify whether patients had moved between  $R_2^*$  and SIR measurements, in which case new ROIs were placed. Ghosting artifacts caused by aortic blood flow were present in SIR measurements before November 2012 (when saturation slabs were added). Separate ROIs were placed to remove these artifacts from the liver and muscle ROIs (Fig. 1E, F). Some reports indicate that susceptibility artifacts may affect  $R_2^*$  measurements when using a single ROI in liver segments VII or VIII [16]. Due to the limited number of slices, we did not formally assess segmental variations of  $R_2$ ,  $R_2^*$ , or  $LIC_{SIR}$  in this study.

The respiratory triggering applied for  $R_2$  data acquisition resulted in slight changes in slice positioning so that new ROIs were placed using  $R_2$  source data as described above.

### $LIC_{SIR}$

The calculations published by Gandon et al. were entered into the aforementioned program [7, 17], which automatically chooses the most reliable SIR (i.e., T1, PD, T2, T2+, or T2++) which is converted to  $LIC_{SIR}$ . The mean  $LIC_{SIR}$  of three slices was used and, when one or more values exceeded the 350  $\mu\text{mol/g}$  threshold, the final value was noted as  $>350 \mu\text{mol/g}$ . In two subanalyses, the  $R_2$  and  $R_2^*$  values and the individual SIR ratios in patients with  $LIC_{SIR} > 350 \mu\text{mol/g}$  were evaluated.



**Fig. 1.** Placement of ROIs. **A–F** The placement of ROIs on the data. **A–C** How the ROIs for the total liver parenchyma (**A**) and intrahepatic vasculature and/or gall bladder (**B**) are drawn and the result of subtraction in (**C**). **D** The

ROI-placement on the paraspinal muscles for SIR calculations. **E, F** The placement of a ghosting artifact ROI (**E**) and the final liver parenchyma ROI (**F**) obtained by subtracting (**E**) from (**C**).

## $R_2^*$

In magnitude images, the noise is distributed in a non-Gaussian manner. This is known as Rician noise [18]. At high signal levels, the non-zero mean has a negligible effect on the average signal, but near the noise level, a noise bias exists which needs to be taken into account when fitting  $R_2^*$ . We explored three different fit routines: a truncated exponential fit (A) [19, 20], an exponential + constant fit (B) [9, 21], and an exponential + Rician noise (C).

The truncated exponential method A is considered the reference standard, but is time-consuming, where methods B + C do not require further manual input. We compared method B and C with method A as reference using Bland–Altman analysis and  $R_2^*$  data from a single reader (EMA). Based on this comparison (mean paired difference ( $\bar{d}$ ) was 0.8 Hz for A–C and 33.6 Hz for A–B), we employed method C (Rician noise bias) for the remaining analyses [22, 23].

$R_2^*$  calculation was thus performed with a monoexponential model (Eq. 1) with a Rician noise factor. In Eq. 1,  $E_R$  describes the Rice distribution (Online Resource 1), where  $\sigma$  is a noise parameter and  $S_0 \times e^{-R_2^* \times TE}$  reflects the true magnitude value. Data were averaged inside the ROI before data fitting (average-then-fit).

$$S(TE) = E_R \cdot (S_0 \cdot e^{-R_2^* \cdot TE}, \sigma) \quad (1)$$

The effect of intrahepatic fat on  $R_2^*$  was assessed by applying a biexponential model in a subset ( $n = 10$ ) with definite presence of fat, as identified by the presence of an oscillating signal intensity decay over time.  $R_2^*$  values with and without correction were compared using Bland–Altman analysis. The ( $\bar{d}$ ) was 0.1 Hz—indicating low overall fat content in this cohort—and deemed negligible compared to the subset mean of 70 Hz. Monoexponentially fitted  $R_2^*$  values were used for all comparisons.

## $R_2$

For  $R_2$  calculation an average-then-fit routine was applied using a biexponential model as shown in Eqs. 2 and 3. In Eq. 2,  $S_T(TE)$  is the signal intensity without noise at time TE,  $S_0$  is the signal intensity at TE = 0, and  $R_2$  is the relaxation rate. The subscripts  $a$  and  $b$  indicate fast and slow relaxation components, respectively. For  $R_2$ , Rician noise bias was approximated by the Pythagorean addition of an extra fit parameter, the noise factor ‘ $v$ ’ in Eq. 3.

$$S_T(TE) = S_{0,a} \cdot e^{-R_{2,a} \cdot TE} + S_{0,b} \cdot e^{-R_{2,b} \cdot TE} \quad (2)$$

$$S(TE) = \sqrt{S_T(TE) + v^2}. \quad (3)$$

In the biexponential model, an iron-dense and an iron-sparse component are assumed, with short and long

$R_2$ , respectively. For further comparisons with  $LIC_{SIR}$  and  $R_2^*$ , the bulk  $R_2$  was calculated (Eq. 4) in accordance with the literature [8, 9, 14].

$$R_2 = \frac{S_{0,a} \cdot R_{2,a} + S_{0,b} \cdot R_{2,b}}{S_{0,a} + S_{0,b}} \quad (4)$$

## Comparison with the literature

The relations between the  $LIC_{SIR}$ ,  $R_2$ , and  $R_2^*$  were compared to published regression analysis results based on either biopsy-proven LIC ( $LIC_{BIOPSY}$ ) [8, 9, 19–21] or  $LIC_{SIR}$  [14].

## Statistical analyses

Data are described as number (%) or median (interquartile range, IQR). Results of observers were compared using a Friedman test and Wilcoxon Signed-Rank test as post hoc. Success rates are defined as the number of correctly acquired scans of at least “adequate” quality divided by the total number of measurements. These were compared using a McNemar test. Correlations were assessed with Spearman’s correlation coefficients ( $r_s$ ), interobserver agreement with two-way random, and absolute intraclass correlation coefficients (ICCs). Both were graded according to Landis et al. [24]. Bland–Altman analysis was performed to compare accuracy between the three MRI methods for a single observer and compare the performance of the three observers [22]. In a separate analysis, the calculated  $R_2$  and  $R_2^*$  values were converted to  $LIC_{R_2(*)}$  values in  $\mu\text{mol/g}$  using the formulas provided by St. Pierre et al. and Garbowski et al. [8, 20] as these were established with image analysis protocols similar to ours.

ROC-analyses were performed for  $LIC_{SIR}$ ,  $R_2$ , and serum values with significant correlation with  $R_2^*$  to establish their diagnostic accuracy to identify increased  $R_2^*$ , i.e.,  $\geq 44$  Hz [9].  $R_2^*$  was chosen as a reference value as it had the best success rate and shortest acquisition time. The optimal cut-off value for  $R_2$  was found by optimizing the Youden index, while for  $LIC_{SIR}$  we used the established cut-off value of  $> 36 \mu\text{mol/g}$ .  $P$  values of  $< 0.05$  were accepted as statistically significant. Statistical analyses were performed using SPSS Version 22 (IBM Corp, Armonk, NY), MedCalc Statistical Software version 16.2.0 (MedCalc Software bvba, Ostend, Belgium; <https://www.medcalc.org>; 2016), and GraphPad Prism 5.0 (GraphPad Software, La Jolla, CA).

## Results

### Patients

Between January 1, 2008 and December 31, 2013, a total of 114 patients (M/F: 74/40) underwent 144 MRI-scans

**Table 2.** Patient characteristics

	Number (%) or median (IQR)
Patients	114
Male/female	74/40 (65/35%)
Age (years)	44 (28.5–58.1)
Indications	
Sickle cell anemia	21/114 (19%)
MDS <sup>a</sup> /leukemia	19/114 (17%)
Thalassemia	17/114 (15%)
Gaucher's disease	16/114 (14%)
Hemochromatosis	14/114 (12%)
Hemosiderosis (not specified)	6/114 (5%)
Other	21/114 (18%)

<sup>a</sup> MDS myelodysplastic syndrome

for routine LIC measurement. Patient characteristics and clinical indications for LIC measurement are described in Table 2. Thirty patients had multiple measurements. To prevent a repeated measurements effect on correlation assessment between  $LIC_{SIR}$ ,  $R_2$ , and  $R_2^*$ , only the 114 baseline measurements were used. SIR,  $R_2$ , and  $R_2^*$  data were available for 108/114 (95%), 72/114 (63%), and 113/114 (99%) baseline measurements.

### MRI success rates

Five SIR measurements were classified unsuccessful because a surface coil was used, one due to erroneous TR/TE combinations. Furthermore, image quality was inadequate (respiration artifacts) in a single patient (only  $R_2$  and  $R_2^*$  acquired). Hence, SIR was successful in 102/114 (89%),  $R_2$  in 71/114 (62%), and  $R_2^*$  in 112/114 (98%) subjects. The success rate of  $R_2$  was lower than that of SIR and  $R_2^*$  ( $P < 0.0001$ , each). Missing datasets were presumed to not have been scanned, with time constraints and respiratory triggering problems as the major cause of the low success rate of the  $R_2$  measurement. For subsequent analyses, only successful baseline measurements were used.

### Interobserver agreement

$LIC_{SIR}$  and  $R_2$  values differed between observer 1 and the other observers (Table 3). However, these differences (median values: 80–85  $\mu\text{mol/g}$  and 33–34 Hz for  $R_2$ ) would be negligible in clinical practice. This was confirmed by high ICCs for SIR,  $R_2$ , and  $R_2^*$  of 0.998, 0.997, and 0.999, respectively. Bland–Altman analysis between pairs of observers showed a single outlier for SIR, while  $R_2$  and  $R_2^*$  showed differences up to 5% for higher values, reflecting the uncertainties in the data fit at very high LIC (Online Resource 1).

### $LIC_{SIR}$ , $R_2$ , and $R_2^*$

Median (IQR)  $LIC_{SIR}$ ,  $R_2$ , and  $R_2^*$  (given for observer 1 and  $LIC_{SIR} < 350 \mu\text{mol/g}$ ) were 84 (30–205), 33 (23–48),

**Table 3.** MRI interobserver agreement: median (IQR) values

MRI method	Observer 1	Observer 2	Observer 3	$P$ value
$LIC_{SIR}$ ( $\mu\text{mol/g}$ )	84 (30–205)	80 (25–197)	85 (26–196)	$< 0.001^a$
$R_2$ (Hz)	33 (23–48)	34 (24–49)	34 (24–49)	$< 0.001^a$
$R_2^*$ (Hz)	123 (56–321)	126 (55–326)	123 (55–317)	0.092

<sup>a</sup> Post hoc analysis using Wilcoxon Signed-Rank tests showed that  $LIC_{SIR}$  and  $R_2$  values of observer 1 differed significantly from either observer 2 or 3 (who did not differ from each other)

and 123 (56–321).  $LIC_{SIR}$  correlated positively with  $R_2$  and  $R_2^*$  with  $r_S$  of 0.90 (95% confidence interval (CI) 0.84–0.94,  $P < 0.0001$ ,  $n = 57$ ) and 0.98 (95% CI 0.97–0.99,  $P < 0.0001$ ,  $n = 87$ ), respectively.  $R_2$  correlated positively with  $R_2^*$ :  $r_S$  of 0.95 (95% CI 0.93–0.97,  $P < 0.0001$ ,  $n = 71$ ). Figure 2A, B shows scatter plots of (SIR-based or biopsy-proven) LIC against  $R_2$  and  $R_2^*$ . Solid lines indicate regression analysis results (95% CI bands as dashed lines). In our patient cohort,  $R_2$  increased linearly with  $LIC_{SIR}$  (Eq. 5), while  $R_2^*$  appeared to have a clear non-linear relationship with  $LIC_{SIR}$ , well described by a quadratic polynomial (Eq. 6).

$$R_2 = 15.5 + 0.107 \cdot LIC_{SIR} \quad (5)$$

$$R_2^* = 42.7 + 0.142 \cdot LIC_{SIR} + 4.02 \times 10^{-3} \cdot LIC_{SIR}^2 \quad (6)$$

The  $LIC_{SIR}$  upper threshold of 350  $\mu\text{mol/g}$  was reached in 15/102 (15%) measurements. In these measurements, only the T1W SIR correlated with  $R_2^*$ , with  $r_S$  of  $-0.72$  (95% CI  $-0.9$  to  $-0.31$ ,  $P = 0.003$ ,  $n = 15$ ). Figure 3 shows the T1W SIR against  $R_2^*$ , indicating that for  $LIC_{SIR} > 350 \mu\text{mol/g}$ , the discriminatory value of the T1W SIR becomes progressively smaller.

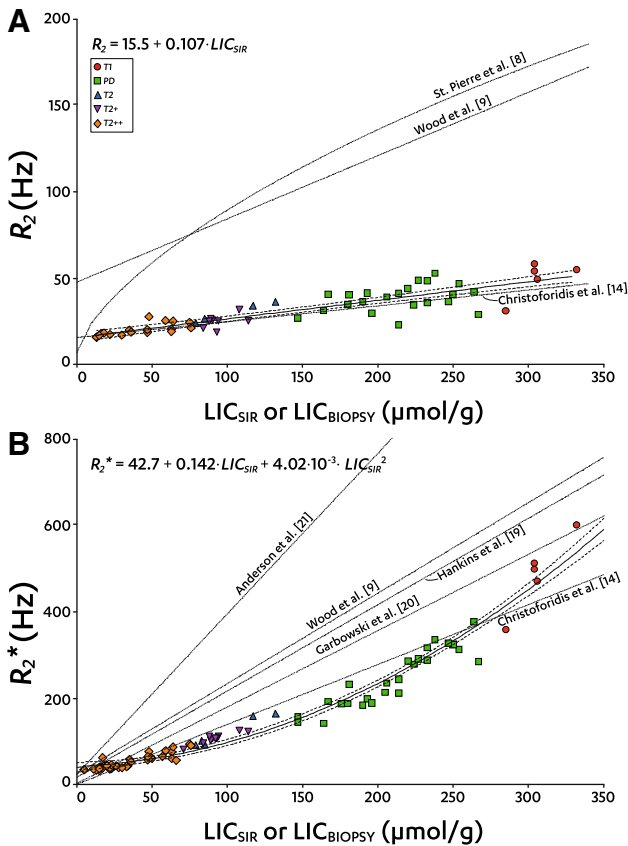
### Comparison with the literature

Figure 2A, B also shows published regression lines between either  $LIC_{SIR}$  or  $LIC_{BIOPSY}$  and  $R_2$  (Fig. 2A) and  $R_2^*$  (Fig. 2B). Contrary to our finding, these lines indicate a linear increase of  $R_2^*$  as LIC increases, and a non-linear increase of  $R_2$  as LIC increases. To assess whether this is caused by  $LIC_{SIR}$  or by  $R_2$  or  $R_2^*$ , we applied established conversion formulae to convert our  $R_2$  (Eq. 7) and  $R_2^*$  (Eq. 8) values to LIC values [8, 20]. We then compared these  $LIC_{R_2^*}$  and  $LIC_{R_2}$  values to our  $LIC_{SIR}$  values.

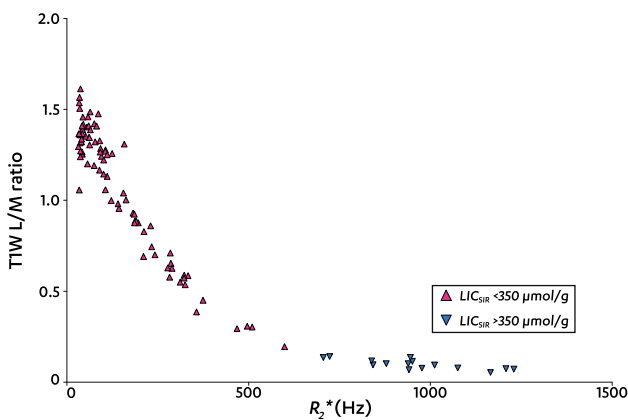
$$LIC_{R_2} (\mu\text{mol/g}) = 17.91 \cdot \left( 29.75 + \sqrt{(900.7 - 2.283 \cdot R_2)} \right)^{1.424} \quad (7)$$

$$LIC_{R_2^*} (\mu\text{mol/g}) = \frac{0.029 \cdot R_2^{*1.014}}{5.585 \cdot 10^{-2}} \quad (8)$$

These established conversion formulae show a non-linear relation between  $R_2$  and true LIC (Eq. 7) and



**Fig. 2.**  $LIC_{SIR}$  or  $LIC_{BIOPSY}$  against  $R_2$  and  $R_2^*$ . **A, B** Scatter plots of  $LIC_{SIR}$  against  $R_2$  (**A, top**) and  $R_2^*$  (**B, bottom**) for all successful baseline measurements. Data points are grouped by SIR LIC type: ● T1; ■ PD; ▲ T2; ▼ T2+; and ◆ T2++. Regression results (equations given in the figures) are shown by *solid lines*, with *dotted* 95% CI bands indicating the goodness of the fit. Additional *dotted regression lines* are based on regression analyses reflecting  $LIC_{BIOPSY}$  [8, 9, 19–21] or  $LIC_{SIR}$  [14].



**Fig. 3.** T1W liver-to-muscle SIR against  $R_2^*$ . This shows a scatter plot of  $R_2^*$  values (x-axis) against the liver-to-muscle SIR (y-axis) of successful baseline T1W SIR measurements. Data are grouped into the following: ▲  $LIC_{SIR} < 350$  and ▼  $LIC_{SIR} > 350$   $\mu\text{mol/g}$ .

linear relation between  $R_2^*$  and true LIC (Eq. 8). Hence, the scatter plot between  $LIC_{R_2^*}$  and  $LIC_{SIR}$  also revealed a quadratic relation, and that between  $LIC_{SIR}$  and  $LIC_{R_2}$  a linear one (data not shown).

### Diagnostic accuracies of $LIC_{SIR}$ , $R_2$ , and serum values

Serum total iron, transferrin, transferrin-saturation, and ferritin were available for 56, 56, 54, and 96 out of 114 measurements. All four correlated significantly with  $R_2^*$ , with best correlation for ferritin at  $r_S = 0.80$  ( $P < 0.0001$ ,  $n = 94$ ).

Increased  $R_2^*$  ( $\geq 44$  Hz) was present in 91 subjects. Of the MRI and serum methods,  $R_2$  and ferritin had best diagnostic accuracies to detect increased  $R_2^*$  (Table 4). Figure 4A–C shows true and false positive and negative results of  $R_2$  (Fig. 4A),  $LIC_{SIR}$  (Fig. 4B), and ferritin (Fig. 4C) for establishing increased  $R_2^*$ .

## Discussion

This study shows that for routine clinical MRI-based LIC measurements SIR and  $R_2^*$  are more often successful than  $R_2$ . Interobserver agreement was near perfect ( $ICC > 0.9$ ) for all methods.  $R_2$  and  $R_2^*$  methods provided relaxation rates when the SIR-threshold ( $> 350$   $\mu\text{mol/g}$ ) was already exceeded. This gives them an advantage over SIR in subjects with transfusional hemosiderosis (at least 55% of our population), when LIC values can easily surpass 350  $\mu\text{mol/g}$ . The combination of high success rate, high interobserver agreement, ability to detect changes in LIC over a wide range of LIC values, and single breath-hold acquisition favors the  $R_2^*$  method for LIC measurement.

In our study, the relationship between  $R_2^*$  and  $LIC_{SIR}$  was quadratic and remained quadratic when  $R_2^*$  was expressed as a LIC value using a previously published (biopsy-proven) conversion formula. Other authors report linear relationships. Given the physics of the  $R_2^*$ –iron relationship, which is basically linear [25], this discrepancy arises either from our  $R_2^*$  acquisition and analysis or from the reference standard. To rule out the former, we compared three fit routines. The exponential + Rician noise factor fit provided identical results in a fraction of the required time to the established and widely applied but labor-intensive method of manual truncation before exponential fitting.

With respect to reference standard, St. Pierre et al. [8], Wood et al. [9], Hankins et al. [19], Garbowski et al. [20], and Anderson et al. [21] all used biopsy-determined  $LIC_{BIOPSY}$  as reference standard, whereas we and Christoforidis et al. [14] used the  $LIC_{SIR}$  according to Gandon. Given the similarity of our MRI protocols, it is unsurprising that Christoforidis' and our data points show considerable overlap. Arguably, their linear rela-

**Table 4.** Diagnostic accuracy values to correctly identify increased  $R_2^*$  ( $\geq 44$  Hz)

	$R_2$	LIC <sub>SIR</sub>	Iron	Transferrin	Transferrin-%	Ferritin
Cases	64/64	75/80	18/41	36/41	20/40	72/80
Cut-off	$\geq 18.3$ Hz	$\geq 36$ mg/g	$\geq 22.6$	$\leq 2.21$	$\geq 0.40$	$\geq 524$
AUROC	1.00 (0.95–1.0)	0.97 (0.91–0.99)	0.66 (0.53–0.79)	0.84 (0.72–0.93)	0.77 (0.64–0.87)	0.98 (0.93–1.0)
Sensitivity	100% (94.4–100%)	93.8% (86.0–97.9%)	43.9% (28.5–60.3%)	87.8% (73.8–95.9%)	50.0% (33.8–66.2%)	90.0% (81.2–95.6%)
Specificity	100% (59.0–100%)	100% (83.9–100%)	100% (76.8–100%)	71.4% (41.9–91.6%)	92.3% (64.0–99.8%)	100% (76.8–100%)
PPV	100% (93.7–100%)	100% (95.2–100%)	100% (82.6–100%)	92.4% (79.8–98.3%)	96.3% (78.3–100%)	100% (94.7–100%)
NPV	100% (77.3–100%)	80.2% (59.7–93.2%)	31.1% (16.7–48.7%)	59.7% (30.3–84.7%)	31.8% (16.4–50.9%)	71.7% (51.0–87.3%)

AUROC area under the ROC curve, PPV positive predictive value, NPV negative predictive value  
Values in parentheses reflect the 95% confidence intervals

tion between LIC<sub>SIR</sub> and  $R_2^*$  could also be described by a quadratic polynomial.

Apart from the linear relationship, the other authors report much steeper increase of  $R_2^*$  as LIC increases [9, 19–21]. Anderson et al.’s very steep increase could be due a long TE1 of 2.2 ms compared to all other studies (range of TE1: 0.8–0.99 ms) that hampers the ability to accurately estimate high  $R_2^*$  values. The fact that the control values of  $R_2^*$  in subjects without iron overload in those studies but also in this paper hover around 40 Hz is a further argument that the observed difference in LIC– $R_2^*$  does not arise from the  $R_2^*$  acquisition or analysis but from the reference standard.

Hence, the most likely cause of the deviating quadratic relation between  $R_2^*$  and estimated LIC is the piecewise sampling of the LIC range with five differently weighted GRE-sequences for LIC<sub>SIR</sub>. This has artificially imposed a quadratic behavior on the actually linear relationship between  $R_2^*$  and true LIC<sub>BIOPSY</sub>. If one looks at the fundamental GRE signal equation (Eq. 9), where PD is proton density and  $\alpha$  is flip angle and applies this to the liver-to-muscle signal intensity ratio, the PD and  $\sin(\alpha)$  terms drop out. By taking the natural logarithm, we find Eqs. 10 and 11. The latter proves that the relationship between  $R_2^*$  and SIR is logarithmic. Indeed, plotting Fig. 3 with a log-scale for the signal intensity ratio on the  $y$ -axis linearized the line (data not shown).

$$S(\text{TE}) = \frac{\text{PD} \cdot \sin(\alpha) \cdot (1 - e^{-\text{TR}/T_1})}{(1 - \cos(\alpha) \cdot e^{-\text{TR}/T_1})} \cdot e^{-R_2^* \cdot \text{TE}} \quad (9)$$

$$\ln\left(\frac{S_{\text{LIVER}}}{S_{\text{MUSCLE}}}\right) = f(\text{TR}, \alpha, T_1) + \text{TE} \cdot (R_{2^*, \text{LIVER}} - R_{2^*, \text{MUSCLE}}) \quad (10)$$

$$R_{2^*, \text{LIVER}} = \frac{\ln\left(\frac{S_{\text{LIVER}}}{S_{\text{MUSCLE}}}\right) - f(\text{TR}, \alpha, T_1)}{\text{TE}} + R_{2^*, \text{MUSCLE}} \quad (11)$$

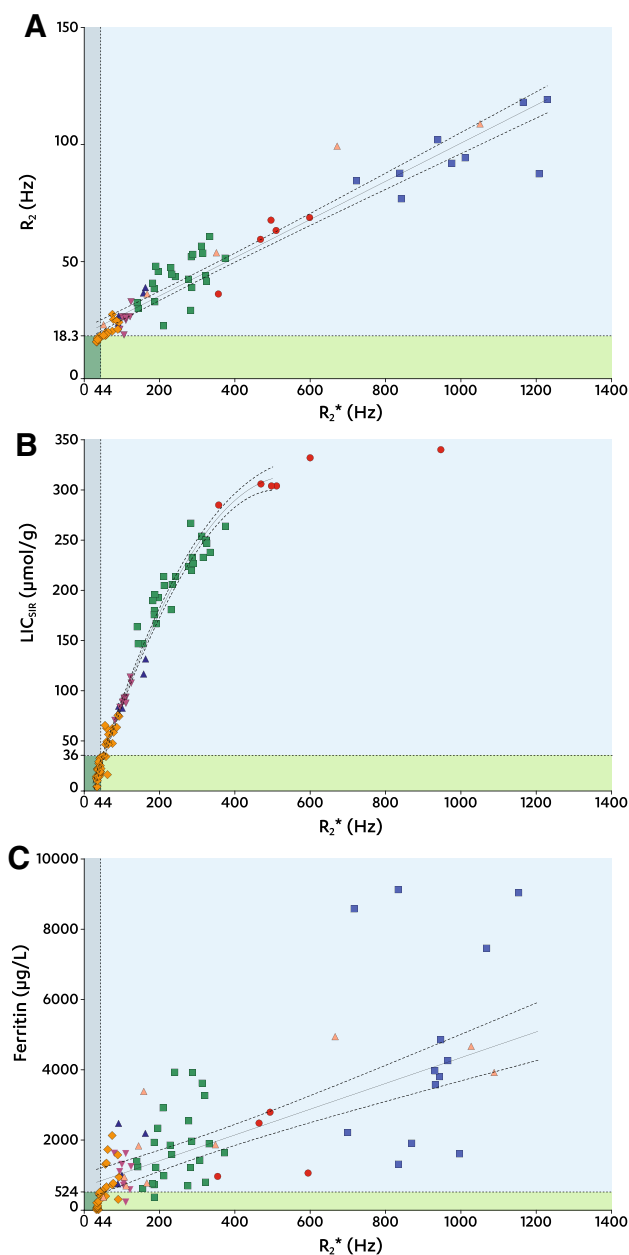
For  $R_2$ , single- and multiecho SE acquisitions are possible: multiecho SE decreases  $R_2$  due to residual signal of stimulated echoes at a given TE. Single-echo SE increases  $R_2$  because long TEs cause increased sensitivity to diffusion, hence increased signal loss at a given TE.

Reported single-echo SE  $R_2$  values [8, 9] were concordantly higher for the same estimated LIC compared to multiecho SE results as in this study and in [14]. In terms of  $R_2$  data fitting, we as many others applied a biexponential model and we did not assess non-exponential decay models as for instance proposed by Jensen et al. [26].

The main limitation of our study is the lack of biopsy confirmation. In our center, liver biopsy for iron determination is seldom performed. Both the national, European and American guidelines recommend reluctance in performing biopsy and underline the high sensitivity of MRI [15, 27, 28]. Moreover, differing processing steps to obtain LIC<sub>BIOPSY</sub> are reported, compromising generalizability. In Gandon’s method, paraffin-embedded liver biopsy specimens are dewaxed using a protocol with a triple xylene wash to remove lipid solids from the sample. This approach was shown to have an elevating effect on the dry weight liver iron calculation compared to processing fresh tissue samples [29]. Another limitation is the fact that we did not perform multipeak fat-correction on complex data [10]. This was not feasible with only magnitude data available. Comparison to other literature is further hampered by the use of different image acquisition and postprocessing protocols which directly influence the calibration curves between the reference standard and the index test. We have opted to compare our findings to calibration curves obtained with similar postprocessing protocols.

ROC-analyses showed that  $R_2$  and ferritin have the highest diagnostic accuracy to identify increased  $R_2^*$  ( $\geq 44$  Hz). Both ferritin ( $\geq 524$   $\mu\text{g/L}$ ) and  $R_2$  ( $\geq 18.3$  Hz) had positive predictive values of 100%, but the wide distribution of ferritin levels for  $R_2^* \geq 44$  Hz indicates that it cannot be used confidently to follow-up treatment nor accurately determine the LIC. In contrast,  $R_2$  shows a different picture with a close distribution around the regression line. In addition, ferritin lacks the spatial information that MRI provides, allowing segmental LIC measurement and follow-up.

$R_2$  datasets were missing (i.e., not scanned) in 42/114 (37%) subjects. As  $R_2$  is part of our routine scan protocol, this illustrates that the long and artifact-prone  $R_2$



**Fig. 4.**  $R_2$ ,  $LIC_{SIR}$ , and ferritin against  $R_2^*$ . **A–C** Scatter plots between  $R_2^*$  ( $x$ -axes) and  $R_2$ ,  $LIC_{SIR}$ , and serum ferritin ( $y$ -axes). Dotted lines at  $x = 44$  and at  $y = 18.3$  (**A**),  $y = 36$  (**B**), and  $y = 524$  (**C**) indicate the thresholds for  $R_2$ ,  $LIC_{SIR}$ , and serum ferritin to identify increased  $R_2^*$  (Table 4). Data points are grouped by SIR LIC type: ● T1; ■ PD; ▲ T2; ▼ T2+; ◆ T2++; ■ >350; and ▲ no  $LIC_{SIR}$  available. Regression results are shown by the solid lines with dotted 95% CI bands indicating the goodness of the fit. Shaded areas indicate true positive (■), true negative (■), false positive (■), and false negative (■), respectively.

series is skipped first by the radiographer. This makes the  $R_2$  series less suited as first choice for LIC measurement.

Our results favor the use of  $R_2^*$  measurements for daily clinical practice with the use of an exponen-

tial + Rician noise fit method to save time in analysis. The recommendation to (only) use  $R_2^*$  comes with cautions. It requires careful consideration of scan parameters which should be kept equal for all measurements. Ideally, routine quality control with phantom testing should be performed.

In conclusion, as  $R_2^*$  can be obtained in a single breath-hold with excellent success rates, high interobserver agreement, and ability to detect changes over a wide range of LIC values and is available from all major vendors without additional per-scan costs, it is our first choice for LIC measurement.

*Acknowledgments.* The authors would like to acknowledge Paul F. Groot for anonymizing the data and Shandra Bipat for providing advice on statistical analyses.

*Compliance with ethical standards*

*Conflict of Interest* The authors declare that they have no conflict of interest.

*Ethical standard/informed consent* This was a retrospective study using data obtained in routine clinical practice that were anonymized before analysis. In light of the respective nature of the study, the obligation to obtain informed consent was waived by the Medical Ethical Committee of the AMC Amsterdam.

*Open Access* This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Tavill AS, AASLD, ACG (2001) Diagnosis and management of 2 hemochromatosis. *Hepatology* 33:1321–1328. doi:10.1053/jhep.2001.24783
2. Pietrangelo A (2003) Haemochromatosis. *Gut* 52(Suppl 2):ii23–ii30. doi:10.1136/gut.52.suppl\_2.ii23
3. Queiroz-Andrade M, Blasbalg R, Ortega CD, et al. (2009) MR imaging findings of iron overload. *Radiographics* 29:1575–1589. doi:10.1148/rg.296095511
4. Bravo AA, Sheth SG, Chopra S (2001) Liver biopsy. *N Engl J Med* 344:495–500. doi:10.1056/NEJM200102153440706
5. Sirlin CB, Reeder SB (2010) Magnetic resonance imaging quantification of liver iron. *Magn Reson Imaging Clin N Am* 18:359–381. doi:10.1016/j.mric.2010.08.014
6. Gandon Y, Olivie D, Guyader D, et al. (2004) Non-invasive assessment of hepatic iron stores by MRI. *Lancet* 363:357–362. doi:10.1016/S0140-6736(04)15436-6
7. Gandon Y. Rennes—hemochromatosis. Y Gandon, Rennes. 10-06-2001. <http://www.radio.univ-rennes1.fr/Sources/EN/Hemo.html>. Accessed October 16, 2015.
8. St Pierre TG, Clark PR, Chua-anusorn W, et al. (2005) Noninvasive measurement and imaging of liver iron concentrations using proton magnetic resonance. *Blood* 105:855–861. doi:10.1182/blood-2004-01-0177
9. Wood JC, Enriquez C, Ghugre N, et al. (2005) MRI  $R_2$  and  $R_2^*$  mapping accurately estimates hepatic iron concentration in transfusion-dependent thalassemia and sickle cell disease patients. *Blood* 106:1460–1465. doi:10.1182/blood-2004-10-3982
10. Hernando D, Kramer JH, Reeder SB (2013) Multiplex fat-corrected complex  $R_2^*$  relaxometry: theory, optimization, and clinical



- validation. *Magn Reson Med* 70:1319–1331. doi:10.1002/mrm.24593
11. Krafft AJ, Loeffler RB, Song R, et al. (2015) Does fat suppression via chemically selective saturation affect  $R_2^*$ -MRI for transfusional iron overload assessment? A clinical evaluation at 1.5T and 3T. *Magn Reson Med*. doi:10.1002/mrm.25868
  12. Yokoo T, Yuan Q, Senegas J, Wiethoff AJ, Pedrosa I (2015) Quantitative  $R_2^*$  MRI of the liver with rician noise models for evaluation of hepatic iron overload: Simulation, phantom, and early clinical experience. *J Magn Reson Imaging* 42:1544–1559. doi:10.1002/jmri.24948
  13. Ibrahim EH, Khalifa AM, Eldaly AK (2016) MRI  $T_2^*$  imaging for assessment of liver iron overload: study of different data analysis approaches. *Acta Radiol*. doi:10.1177/0284185116628337
  14. Christoforidis A, Perifanis V, Spanos G, et al. (2009) MRI assessment of liver iron content in thalassamic patients with three different protocols: comparisons and correlations. *Eur J Haematol* 82:388–392. doi:10.1111/j.1600-0609.2009.01223.x
  15. Swinkels DW, van Bokhoven MA, Castel A, et al. (2007) *Richtlijn Diagnostiek en behandeling van hereditaire hemochromatose*. Utrecht: Nederlandse Internisten Vereniging en Nederlandse Vereniging voor Klinische Chemie
  16. Meloni A, Luciani A, Positano V, et al. (2011) Single region of interest versus multislice  $T_2^*$  MRI approach for the quantification of hepatic iron overload. *J Magn Reson Imaging* 33(2):348–355. doi:10.1002/jmri.22417
  17. Gandon Y. Gandon calculations. Y Gandon, Rennes. 10-06-2001. <http://www.radio.univ-rennes1.fr/Images/Externe15.js>. Accessed October 16, 2015
  18. Gudbjartsson H, Patz S (1995) The Rician distribution of noisy MRI data. *Magn Reson Med* 34:910–914
  19. Hankins JS, McCarville MB, Loeffler RB, et al. (2009)  $R_2^*$  magnetic resonance imaging of the liver in patients with iron overload. *Blood* 113:4853–4855. doi:10.1182/blood-2008-12-191643
  20. Garbowski MW, Carpenter JP, Smith G, et al. (2014) Biopsy-based calibration of  $T_2^*$  magnetic resonance for estimation of liver iron concentration and comparison with  $R_2$  Ferriscan. *J Cardiovasc Magn Reson*. doi:10.1186/1532-429X-16-40
  21. Anderson LJ, Holden S, Davis B, et al. (2001) Cardiovascular  $T_2$ -star ( $T_2^*$ ) magnetic resonance for the early diagnosis of myocardial iron overload. *Eur Heart J* 22:2171–2179. doi:10.1053/ehj.2001.2822
  22. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–310. doi:10.1016/S0140-6736(86)90837-8
  23. Akkerman EM, Runge JH, Troelstra MA, Nederveen AJ, Stoker J (2015) Non-linear relationship between estimated liver iron concentration and  $R_2^*$ . *ISMRM* 3268
  24. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
  25. Ghugre NR, Wood JC (2011) Relaxivity-iron calibration in hepatic iron overload: probing underlying biophysical mechanisms using a Monte Carlo model. *Magn Reson Med* 65:837–847. doi:10.1002/mrm.22657
  26. Jensen JH, Chandra R (2002) Theory of nonexponential NMR signal decay in liver with iron overload or superparamagnetic iron oxide particles. *Magn Reson Med* 47:1131–1138. doi:10.1002/mrm.10170
  27. European Association for the Study of the Liver (2010) EASL clinical practice guidelines for HFE hemochromatosis. *J Hepatol* 53:3–22. doi:10.1016/j.jhep.2010.03.001
  28. Bacon BR, Adams PC, Kowdley KV, Powell LW, Tavill AS (2011) Diagnosis and management of hemochromatosis: 2011 practice guideline by the American Association for the Study of liver diseases. *Hepatology* 54:328–343. doi:10.1002/hep.24330
  29. Butensky E, Fischer R, Hudes M, et al. (2005) Variability in hepatic iron concentration in percutaneous needle biopsy specimens from patients with transfusional hemosiderosis. *Am J Clin Pathol* 123:146–152. doi:10.1309/PUUXEGXDLH26NXA2