

# MRI R2 and R2\* mapping accurately estimates hepatic iron concentration in transfusion-dependent thalassemia and sickle cell disease patients

John C. Wood, Cathleen Enriquez, Nilesh Ghugre, J. Michael Tyzka, Susan Carson, Marvin D. Nelson, and Thomas D. Coates

Measurements of hepatic iron concentration (HIC) are important predictors of transfusional iron burden and long-term outcome in patients with transfusion-dependent anemias. The goal of this work was to develop a readily available, noninvasive method for clinical HIC measurement. The relaxation rates R2 (1/T2) and R2\* (1/T2\*) measured by magnetic resonance imaging (MRI) have different advantages for HIC estimation. This article com-

pares noninvasive iron estimates using both optimized R2 and R2\* methods in 102 patients with iron overload and 13 controls. In the iron-overloaded group, 22 patients had concurrent liver biopsy. R2 and R2\* correlated closely with HIC ( $r^2 \geq .95$ ) for HICs between 1.33 and 32.9 mg/g, but R2 had a curvilinear relationship to HIC. Of importance, the R2 calibration curve was similar to the curve generated by other researchers, despite

significant differences in technique and instrumentation. Combined R2 and R2\* measurements did not yield more accurate results than either alone. Both R2 and R2\* can accurately measure hepatic iron concentration throughout the clinically relevant range of HIC with appropriate MRI acquisition techniques. (Blood. 2005;106:1460-1465)

© 2005 by The American Society of Hematology

## Introduction

Hepatic iron concentration (HIC) is used as a surrogate for total iron balance to guide chelation therapy in transfusion-dependent patients.<sup>1,2</sup> Unfortunately, liver biopsy is invasive and provides only indirect information regarding other organ systems. Biomagnetic susceptibility measurements by superconducting quantum interference device, or SQUID, have been used as surrogates by some, but only 4 such devices are currently operating worldwide.<sup>3-7</sup> The use of magnetic resonance imaging (MRI) relaxation time techniques to estimate liver iron concentration has been studied for nearly 20 years.<sup>8-15</sup> Iron shortens T1, T2, and T2\* relaxation times measured by MRI, darkening images in the presence of iron. Since MRI is ubiquitous, it offers great potential for widely accessible, noninvasive estimation of HIC.<sup>16</sup>

The reciprocals of T2 and T2\*, known as R2 and R2\*, are directly proportional to iron and demonstrate the most promising results. Most investigators have described a linear rise in R2 with iron<sup>8-15</sup>; however, these studies have been criticized for their small size, limited dynamic range, and interstudy calibration variability.<sup>16,17</sup> In a study of more than 100 patients, St Pierre et al found a curvilinear R2 relationship between R2 and biopsy HIC over the entire clinically relevant range.<sup>18</sup> More importantly, that study demonstrated measurement stability between multiple imaging platforms. R2 changes have also been qualitatively associated with cardiac iron deposition.<sup>9,14</sup>

There are fewer studies of R2\* for iron quantitation. Anderson et al<sup>19</sup> demonstrated a negative log-linear correlation between liver T2\* and HIC of 0.93 in nonfibrotic livers. The slope of this relationship was  $-1.07$ , predicting a nearly linear rise of R2\* with iron. They also demonstrated clinical correlation between cardiac T2\* measurements and cardiac function.<sup>19</sup> R2 and R2\* methods

theoretically have different sensitivities to subcellular iron distribution, such as might be seen at different iron loads and in cirrhosis or other liver diseases.<sup>19-23</sup> Some investigators have even proposed that the difference between R2\* and R2, called R2', may be a more specific marker of tissue iron.<sup>24,25</sup> To our knowledge, there has been no systematic simultaneous comparison of liver R2 and R2\* methods in identical patients.

Using MRI settings optimized for liver iron estimation, we compared the relationship of R2 and R2\* values to one another in 102 patients with iron overload and 13 healthy controls. MRI was validated to biopsy-measured iron concentration in 22 of the iron overload patients having a liver iron load ranging from 1.3 mg/g to 57.8 mg/g (dry weight).

## Patients, materials, and methods

### Patient population

One hundred and two patients underwent a total of 132 comprehensive iron evaluations between August 2002 and August 2004. Patients were primarily referred for cardiac T2\* and cardiac function analysis but consented to MRI HIC assessment as well. The study was approved by the institutional review board at Children's Hospital Los Angeles. Informed consent was provided according to the Declaration of Helsinki. Iron overload was assessed in patients with thalassemia major ( $n = 57$ ), sickle cell disease ( $n = 34$ ), thalassemia intermedia ( $n = 6$ ), aplastic anemia ( $n = 3$ ), hemochromatosis ( $n = 1$ ), and heme-metabolism defect ( $n = 1$ ). Twenty two (of the 102) patients were scheduled to have their MRI examinations during or immediately following a clinically indicated liver biopsy. Mean time

From the Department of Pediatrics and the Department of Radiology, Children's Hospital of Los Angeles, CA; and the Department of Biology, California Institute of Technology, Pasadena, CA.

Submitted October 18, 2004; accepted April 15, 2005. Prepublished online as Blood First Edition Paper, April 28, 2005; DOI 10.1182/blood-2004-10-3982.

Supported by the National Heart, Lung and Blood Institute (1R01 HL75592-01A1) and the National Center for Research Resources (General Clinical Research Center [GCRC] RR00043-43) at the National Institutes of Health as well as Novartis Pharma, the Whitaker Foundation, and the

Department of Pediatrics at Children's Hospital of Los Angeles.

**Reprints:** Dr John Wood, Division of Cardiology, Mailstop 34, Children's Hospital Los Angeles, 4650 Sunset Blvd, Los Angeles, CA 90027-0034; e-mail: jwood@chla.usc.edu.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 U.S.C. section 1734.

© 2005 by The American Society of Hematology

between biopsy and MRI examination was  $4.3 \pm 9.5$  days (range, 0-32 days). Two patients had repeat biopsies in the study interval. Biopsy indications were thalassemia major ( $n = 9$ ), sickle cell disease ( $n = 10$ ), thalassemia intermedia ( $n = 2$ ), and Blackfan-Diamond syndrome ( $n = 1$ ).

Liver biopsy and iron quantitation were performed according to standard clinical practice (Mayo Medical Laboratory, Rochester, MN). One complete core was sent fresh in a trace element-free container. Sample wet weight was obtained in 4 patients and was  $5.4 \pm 0.6$  mg (range, 5.1-6.3 mg).

Liver iron concentrations in the patient population were quite high. We were unable to obtain HIC estimates from patients with nontransfusional iron overload, as was done in a previous study,<sup>26</sup> because of ethical and institutional restrictions. Therefore, to better define calibration curves at low iron concentrations, liver R2 and R2\* were collected from 13 healthy volunteers (9 male, 4 female), ages  $29.3 \pm 12.3$  years (range, 12-50 years). Since liver biopsy was not performed in these subjects, it was necessary to estimate HIC from population norms. The upper limit of normal (95%) for HIC is  $55.8 \times \text{age}$  (units of  $\mu\text{g/g}$  dry weight; John Butz, Director, Mayo Medical Laboratory, Rochester, MN). We used an estimated coefficient of variation of 20% to translate the 95% confidence interval to an expected mean value, that is,  $[\text{Fe}]^\mu = [\text{Fe}]^{95\%}/(1 + 1.96 [\text{COV}])$ , with COV indicating coefficient of variation. This yielded an age-matched mean iron estimate of 1.17 mg/g dry weight for our healthy control population.

### MRI techniques

MRI measurements were performed using a 4-element torso coil on a 1.5 T General Electric CVi scanner (General Electric Medical Systems, Milwaukee, WI). Liver R2\* was measured from a single midhepatic slice using a single echo, gradient echo sequence; echo time (TE) was automatically stepped at 0.25-millisecond intervals from 0.8 to 4.8 milliseconds in a single breath-hold. Other imaging parameters included a field of view of  $48 \times 24$ , a flip angle of  $20^\circ$ , a repetition time (TR) of 25 milliseconds, a matrix of  $64 \times 64$ , a slice thickness of 15 mm, and a bandwidth of 83 kHz. Two seconds of dummy scans were performed to achieve longitudinal steady-state prior to data acquisition. Liver R2 was measured from 4 slices using a single echo,  $120^\circ$ - $120^\circ$  Hahn echo, using TEs of 3.5, 5, 8, 12, 18, and 30 milliseconds. A Hahn echo yielded significantly shorter echo times than a conventional  $90^\circ$ - $180^\circ$  pulse combination; this allowed iron characterization over a much greater range. A single TE was acquired per 15-second breath-hold. Other imaging parameters included a field of view of  $48 \times 24$ , a TR of 300 milliseconds, a matrix of  $64 \times 64$ , a slice thickness of 15 mm, a gap of 5 mm, and a bandwidth of 32 kHz. One patient did not have R2 measurements because of technical difficulties with the MRI scanner.

The gradient echo (R2\*) and spin-echo (R2) images were fit to monoexponential equations with a variable offset:  $S(\text{TE}) = Ae^{-\text{TE} \cdot R2^*} + C$  (equation 1).

The constant, C, was necessary to compensate for contributions from instrumentation noise and effects from iron-poor species such as blood and bile. Equation no. 1 was fit to every pixel in the image. A region of interest was drawn around the entire liver boundary, excluding obvious hilar vessels. Mean, standard deviation, and standard error were calculated from all pixels within the region of interest.

### Calibration curves

Linear calibrations between R2\*, R2, R2', and iron were estimated using univariate regression. Regression analysis yielded linear calibrations of the following form:  $[\text{Fe}]_{R2^*} = .0254 \times R2^* + 0.202$  (equation 2),  $[\text{Fe}]_{R2-L} = 0.148 \times R2 - 6.51$  (equation 3), and  $[\text{Fe}]_{R2'} = 0.0329 \times R2'$  (equation 4), where R2' is  $R2^* - R2$ ,  $[\text{Fe}]_{R2^*}$  is the HIC estimated from R2\*,  $[\text{Fe}]_{R2-L}$  is the HIC estimated from R2 using a linear fit, and  $[\text{Fe}]_{R2'}$  is the HIC estimated from R2'.

To better characterize the nonlinear relationship of R2 versus iron, R2 was plotted against  $[\text{Fe}]_{R2^*}$  for all 132 iron examinations (102 patients). This relationship was curvilinear but was well described by the nonlinear calibration proposed by St Pierre et al,<sup>18</sup> given by the following:  $R2 = 6.88 + 26.06[\text{Fe}]^{0.701} - 0.438[\text{Fe}]^{1.402}$  (equation 5).

Estimates of liver iron by the St Pierre et al<sup>18</sup> R2 calibration,  $[\text{Fe}]_{R2-SP}$ , can be achieved from equation no. 5 by completing the square and algebraic manipulations to yield the following (equation 6):

$$[\text{Fe}]_{R2-SP} = (29.75 - \sqrt{900.7 - 2.283R2})^{1.424}$$

Agreement between all calibrations and HIC was assessed using 95% prediction intervals from the linear regression and by Bland-Altman analysis. The Bland-Altman statistic, formed by the difference of 2 measurements divided by the mean of 2 estimates, characterizes both systematic differences (bias) and random fluctuations (variance). Two-sample *t* test was used to decide whether bias between measurement methods was significant. Confidence interval widths for the different MRI methods were compared with one another using a 2-sample variance test. To preserve statistical independence, repeat MRI examinations were excluded from all statistical calculations (regression and Bland-Altman analysis), although values were included in the data graphs.

### Reproducibility

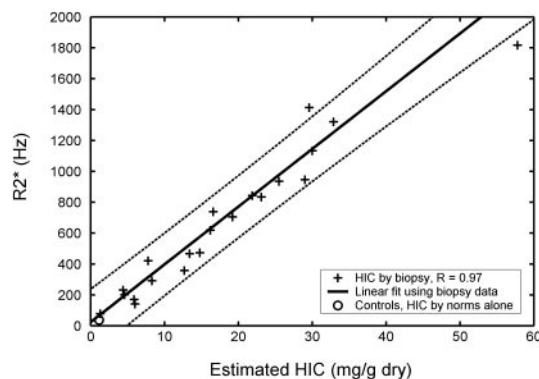
R2 and R2\* measurement reproducibility was assessed in 9 iron-overloaded patients and 3 controls scanned 1 to 3 weeks apart. Identical examinations were performed. Slice-to-slice variability was examined in 4 adjacent slices. For R2\*, this required 3 additional breath-holds and was performed in 5 patients. The R2 method generates 4 slices with the routine examination, so slice-to-slice variability could be generated in every biopsied patient ( $n = 22$ ).

## Results

Liver biopsies were performed without complication. One biopsy specimen for iron quantitation was rejected from analysis prior to iron quantitation because of poor specimen quality. Pathologic analysis demonstrated mild, periportal fibrous expansion in 11 patients and minimal fibrosis in the remainder. Biopsy-measured HIC was uniformly distributed from 1.3 to 32.9 mg/g dry weight except for 1 patient with an HIC of 57.8 mg/g dry weight. This point was a significant outlier with respect to iron concentration and MRI values, and was excluded from statistical analysis.

MRI-estimated iron concentration in the 102 iron-overloaded patients (using average HIC calculated from equations 2 and 6) was  $13.1 \pm 12.3$  mg/g dry weight (range, 1.2-57.3 mg/g dry weight). The patients who underwent liver biopsy tended to have higher MRI HICs:  $17.7 \pm 12.1$  mg/g dry weight (range, 2.1-46.4 mg/g dry weight). Although not statistically different, this likely represents referral bias from physicians regarding the urgency of liver biopsy.

Figure 1 demonstrates R2\* as a function of biopsied HIC. The highest iron concentration is an outlier, but linear agreement



**Figure 1.** Plot of transverse relaxivity R2\* ( $1/T2^*$ ) versus biopsied hepatic iron concentration (HIC) in 21 patients (23 biopsies). R2\* has units of hertz and HIC has units of milligram per gram dry weight of liver. R value was 0.97, and dashed lines indicate 95% prediction intervals for the regression. Average R2\* value for 13 healthy controls is shown for comparison  $\circ$ , plotted using an HIC value estimated from normative data (no biopsy). Repeat MRI and biopsy examinations as well as control data were excluded from statistical calculations.

**Table 1. Comparison of Bland-Altman statistics for different R2 and R2\* methods**

Method	Bias, %	Standard deviation, %	P value vs St Pierre et al <sup>18</sup>	95% confidence interval, %
R2*	1	23	.21	−46–44
R2 linear	−4	26	.43	−55–46
R2 nonlinear	−6	20*	.08	−46–34
R2'	−8	23	.24	−54–38
Average (R2, R2*)	−3	20	.08	−43–37
R2 St Pierre et al <sup>18</sup>	−3	27	—	−56–50

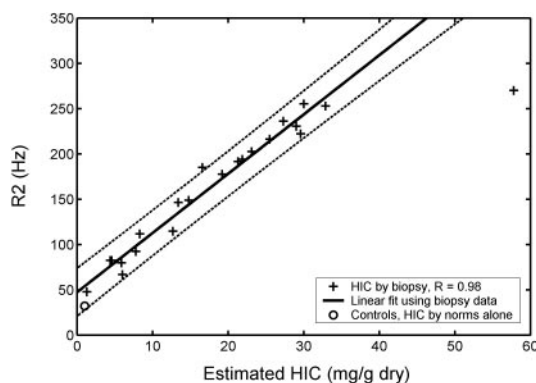
— indicates no comparison possible.

\*P = .15 versus R2 linear.

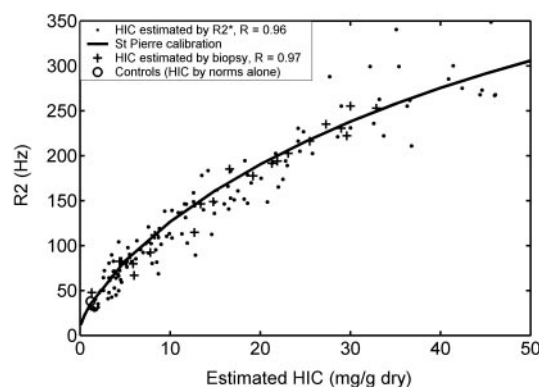
between R2\* and HIC was excellent from 1.3 to 32.9 mg/g dry weight. Regression analysis yielded a correlation coefficient of 0.97, a slope of 37.4 Hz per mg/g dry weight, and a y-intercept of 23.7 Hz. Healthy controls had a mean R2\* of 39.9 Hz  $\pm$  2.8 (SEM); this was plotted against their estimated HICs (1.17 mg/g dry weight). The R2\* fit passes near this point, suggesting reasonable extrapolation to very low liver iron levels. Bland-Altman agreement of  $[\text{Fe}]_{\text{R2*}}$  and biopsy-derived HICs is shown in Table 1. There was no significant bias, and confidence intervals were −46% to 44%, comparable with a recently published R2 methodology.<sup>18</sup>

Figure 2 demonstrates the corresponding relationship between R2 and liver iron concentration. The R2 value for the liver biopsy value of 57.8 was a significant outlier, but the R2-iron relationship appeared close to linear up to 32.9 mg/g. Linear regression between R2 and iron demonstrate a slope of 6.54 Hz per mg/g dry weight, a y-intercept of 47.4 Hz, and a correlation coefficient of 0.98. Limits of agreement between  $[\text{Fe}]_{\text{R2-L}}$  demonstrates statistically insignificant bias (−4%,  $P = .48$ ) and comparable 95% confidence intervals (−55%–46%,  $P = .40$ ) to the work of St Pierre et al.<sup>18</sup> Healthy controls had a mean R2 of 38.2 Hz  $\pm$  1.4 (SEM), plotted again at an estimated HIC of 1.17 mg/g dry weight. Notice that a linear R2 versus iron relationship does not extrapolate well to the healthy controls (55.1 Hz compared with 38.2 Hz,  $P < .001$ ).

Further evidence that the R2 calibration curve is nonlinear comes by plotting the R2-iron relationship for all 132 iron examinations. Figure 3 demonstrates R2 as a function of HIC, measured by biopsy (+ signs) and by  $[\text{Fe}]_{\text{R2*}}$  (solid dots). With the additional examinations, the nonlinear trend is quite obvious. Curvature is pronounced only for HICs less than 7 mg/g, but the



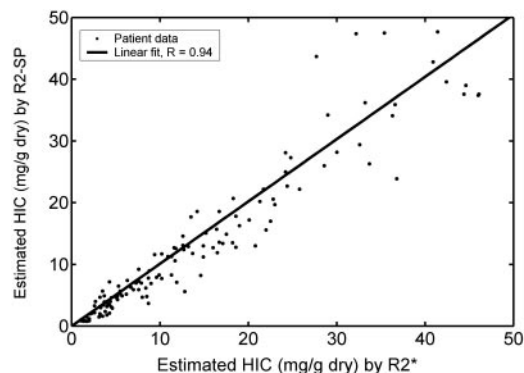
**Figure 2. Plot of transverse relaxivity R2 (1/T2) versus biopsy-measured HIC in 20 patients (22 biopsies).** R value was 0.98, and dotted lines indicate 95% prediction intervals for the regression. Average R2 value for 13 healthy controls is shown by o, plotted using an HIC value estimated from normative data (no biopsy). Repeat MRI and biopsy examinations as well as control data were excluded from statistical calculations.



**Figure 3. R2 versus HIC estimated by biopsy and by R2\*.** R2 versus iron (Figure 2) has been replotted to include all 102 iron overloaded patients (132 examinations). HIC was estimated by R2\* (solid dots) in all examinations and by biopsy (+ signs) in 20 patients (22 examinations). Open circle represents mean R2 from 13 control subjects, plotted using an HIC value estimated from normative data (no biopsy). R2 follows a curvilinear relationship with HIC that is continuous with the mean value observed in non-iron-overloaded subjects. Solid line denotes calibration curve empirically derived by St Pierre et al.<sup>18</sup> Agreement between this curve and estimated HIC was excellent for both biopsy-estimated iron ( $R = 0.97$ ) and R2\*-estimated iron ( $R = 0.96$ ). All repeat MRI and biopsy examinations as well as control data were excluded from statistical calculations.

trend passes through the healthy control estimates. The bold line in Figure 3 represents the calibration curve proposed by St Pierre et al.<sup>18</sup> (equation 5). This curve fits the low-iron behavior well, with a correlation coefficient of 0.97 with respect to biopsy HIC and 0.96 with respect to HIC by R2\*. The limits of agreement between  $[\text{Fe}]_{\text{R2-SP}}$  and  $[\text{Fe}]_{\text{biopsy}}$  were smaller than for a linear calibration, −46% to 34% versus −55% to 46%, although this difference was not statistically significant ( $P = .15$ ). Linear fit to the R2\*-estimated HIC produces comparable R value to the nonlinear fit (0.95) but still has a y-intercept (50.2 Hz) that badly overestimates R2 values in healthy controls.

Agreement between  $[\text{Fe}]_{\text{R2*}}$  (equation 2) and  $[\text{Fe}]_{\text{R2-SP}}$  (equation 6) is demonstrated by scattergram in Figure 4, along with its regression line. Both  $[\text{Fe}]_{\text{R2*}}$  and  $[\text{Fe}]_{\text{R2-SP}}$  estimate rise at the same rate (slope =  $1.01 \pm 0.02$ ), and the correlation between them is 0.94. Despite this,  $[\text{Fe}]_{\text{R2-SP}}$  exhibited a −11% bias toward  $[\text{Fe}]_{\text{R2*}}$  and confidence intervals were broad (−66%–43%). The bias is greatest between HICs of 7 and 25 mg/g dry weight, suggesting that flatter calibration curvature would yield a better fit to our data in this region. In general, variability increased with estimated HIC but is notably larger for HICs more than 30 mg/g.



**Figure 4. Comparison of iron concentration estimated by R2 (equation 6) and by R2\* (equation 2).** Regression slope is  $1.01 \pm 0.02$ , with a correlation coefficient of 0.94. Despite this, the HIC by R2 has an 11% bias relative to values predicted by R2\*, and limits of agreement are broader than for corresponding comparison with biopsy.

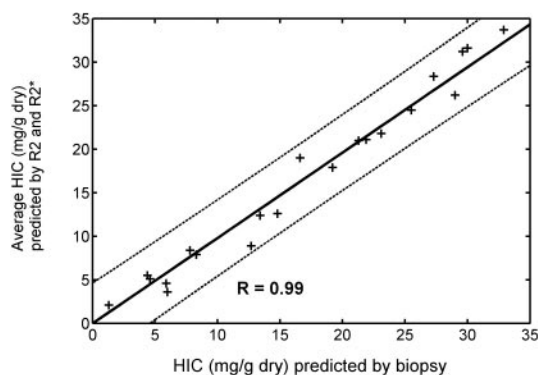


Since HIC estimates by R2 and R2\* measurements exhibit greater disagreement with one another than with biopsy, we examined whether averaging these measurements would improve HIC estimation relative to either technique alone. Figure 5 demonstrates predicted HIC calculated from the average of equations 2 and 6 compared with biopsied HIC values. Correlation coefficient is 0.99 with a slope of 0.98. However, neither 95% prediction intervals nor Bland-Altman limits of agreement were significantly improved relative to either individual measurement (Table 1).

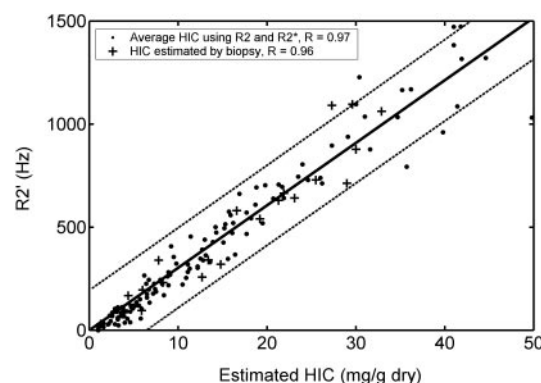
An alternative means to combine R2\* and R2 measurements is to calculate their difference ( $R2^* - R2$ ), also known as  $R2'$ . Figure 6 demonstrates  $R2'$  as a function of HIC estimated by biopsy (+ signs) and by the average of R2\* and R2 iron estimates (solid dots).  $R2'$  rises linearly with HIC, with a correlation coefficient of 0.97 when HIC was estimated by MRI and 0.96 when measured by biopsy. Observed  $R2'$  is less than predicted for a linear relationship when HIC is less than 7 mg/g, corresponding to the greatest nonlinearity in R2 measurements. Bland-Altman agreement of  $[Fe]_{R2'}$  (equation 3) with biopsy values is comparable with R2 and R2\* measurements alone; again, combined measurement yielded no improvement.

A practical assessment of MRI efficacy is illustrated in Table 2. Table 2 compares estimated liver irons by HIC, R2, R2\*,  $R2'$ , and average (R2, R2\*). Using the biopsy HIC value, patients were stratified into 4 classes according to the algorithm of Olivieri and Brittenham<sup>2</sup>: (1) HIC less than 3.2 mg/g indicates concerns for chelator toxicity; (2) HIC of 3.2 to 7.0 mg/g, optimal chelation range; (3) HIC of 7 to 15 mg/g, elevated hepatic iron levels; and (4) HIC more than 15 mg/g, markedly increased iron levels and potential cardiotoxicity. For this population, which was somewhat skewed toward heavy iron overload, there was only one classification "error" with a patient having an HIC by biopsy of 7.8 mg/g classified by R2 as being within the "optimal" range. Thus, in this relatively small cohort, MRI would have generated nearly identical therapeutic decisions as liver biopsy.

R2 and R2\* estimates varied between imaging slices in any given patient. This variability was larger in the R2\* measurements, having a mean coefficient of variation of 7.8%, compared with 4.6% for R2. Both techniques were quite reproducible from exam to exam. Paired R2\* measurements demonstrated a mean difference of 4.0% and a standard deviation of 8.3%. R2 measurements had a mean difference of -0.6% and a standard deviation of 7.4%. The patient subset studied for reproducibility had an MRI-estimated HIC of  $15.9 \pm 13.9$  mg/g dry weight (range, 1.5–41 mg/g



**Figure 5. Average HIC predicted by R2 (equation 6) and R2\* (equation 2) is plotted versus HIC by biopsy in 19 patients.** (Patient with HIC of 57.8 was excluded from the graph and statistics.) + indicates individual data point; solid line, best fit by linear regression; and dotted lines, 95% confidence intervals for the prediction. Regression slope is 0.98, and R value is 0.99. However, Bland-Altman limits of agreement are no better than from the individual measurements (Table 1).



**Figure 6. Plot of  $R2'$  versus HIC.** HIC was estimated by MRI (●) and by biopsy (+). Solid line represents best fit by linear regression, and dotted lines indicate 95% confidence intervals for the prediction.  $R2'$  rises linearly with HIC at iron concentrations greater than 7 mg/g but deviates from linearity at lower iron concentrations. R value is 0.97 with respect to MRI HIC estimates and 0.96 with respect to biopsy HIC estimates. Limits of agreement of  $R2'$  with biopsy are comparable with HIC measurements by R2\* alone (Table 1).

dry weight, which is comparable to the iron burden observed in the general population).

## Discussion

R2 and R2\* methods have theoretical advantages and disadvantages compared with one another. R2 techniques are insensitive to the size and shape of the imaging "voxel" as well as external magnetic inhomogeneities, such as metal clips and air interfaces,

**Table 2. Hepatic iron concentration**

Patient by concentration category	Biopsy	R2*	R2	R2'	R2-R2*
<b>Low</b>					
17	1.3	2.2	2.0	1.0	2.1
<b>Optimal</b>					
11	4.4	6.1	3.5	5.5	5.5
6	4.6	5.3	4.7	4.1	5.1
10	5.9	4.5	4.9	3.1	4.6
9	6.0	3.8	3.5	6.4	3.6
<b>Increased</b>					
20*	7.8	10.9	5.9	11.2	8.4
13	8.3	7.6	8.1	7.6	7.9
19	12.7	9.3	8.4	8.5	8.9
1	13.4	12.1	12.7	11.2	12.4
15	14.8	12.2	13.0	10.5	12.6
<b>High</b>					
12	16.6	19.0	18.9	19.1	19.0
21	19.2	18.1	17.6	17.8	17.9
2b	21.3	20.4	20.2	20.7	22.7
7	21.9	21.6	20.6	21.8	21.1
4	23.1	21.4	22.2	21.1	21.8
2a	25.5	24.0	25.0	24.0	24.5
16b	27.3	32.5	29.4	35.8	31.0
18	29.0	24.2	28.1	23.4	26.2
16a	29.6	36.1	26.3	36.0	31.2
3	30.0	29.0	34.2	28.8	31.6
5	32.9	33.8	33.5	34.9	33.7
8	57.8	46.0	38.2	50.3	42.3
14	16.2	15.9	—	—	—

— indicates a missed R2 measurement due to technical problems.

\*This patient had one concentration that fell into the optimal, rather than increased, concentration range; that value is indicated by underlined italics.

while  $R2^*$  methods can be influenced by these factors. In contrast,  $R2^*$  measurements are more robust to variations in the length scale of iron deposition (ie, they more accurately reflect bulk magnetic susceptibility of tissues).<sup>20,27,28</sup>  $R2^*$  measurements can also be performed in a single breath-hold, while  $R2$  methods take 5 to 20 minutes (depending on technique). However, in this paper, we demonstrate that *both* techniques produce comparable, clinically useful, noninvasive estimates of HIC. Limits of agreement for  $[Fe]_{R2}$  of  $-46\%$  to  $34\%$  and  $[Fe]_{R2^*}$  of  $-46\%$  to  $44\%$  compare favorably with the limits of agreement of  $-56\%$  to  $50\%$  found by St Pierre et al<sup>18</sup> for their  $R2$  method. Despite improved agreement, we do not claim technical superiority, only comparability. Our study had many advantages that would tend to improve both liver biopsy and MRI accuracy, for example, inclusion of younger patients; absence of hepatitis C or significant liver fibrosis in patients; the use of an entire, fresh liver core for assessment; and performance of examinations at a single center. We also had a proportionally greater number of patients with liver irons more than 10 mg/g in which our MRI techniques were quite accurate. In fact, similarities between the 2 studies are far more striking than the differences. Despite using different magnets, different MRI pulse sequences, and different fitting algorithms, we independently generated data consistent with St Pierre et al's<sup>18</sup> nonlinear calibration curve (Figure 3). These results reinforce the portability and reproducibility of  $R2$  techniques if proper care is taken in data collection and analysis. Although our data suggest a slightly flatter calibration curvature for HICs between 7 and 25 mg/g, the behavior at both low and high extremes is quite consistent.

So why have some investigators found linear  $R2$ -HIC relationships and others have found nonlinear calibrations? The most likely explanation lies in inspection of Figures 2 and 3. Curvilinear relationships (in the presence of measurement error) are very difficult to demonstrate unless they are examined over a large range of values (iron concentrations) and large numbers of patients. In Figure 2, the  $R2$ -iron relationship appears well described by a line; the only problem with this fit is its poor extrapolation to low iron. However, when the same relationship is viewed over 132 examinations spanning HICs of 1 to 50 mg/g, the curvilinearity is obvious. Other factors could also play a role in the observed interstudy variation. Some were performed at different magnetic field strengths; this has a profound effect on the magnitude and shape of the calibration curve.<sup>29</sup>  $R2$  measurements using a train of echoes (also known as Carr-Purcell-Meiboom-Gill, or CPMG, sequences) will be lower than those performed by single spin-echo techniques and will vary with echo-spacing.<sup>30</sup> Finally, choice of fitting algorithm can impact the estimated  $R2$  or  $R2^*$ .<sup>31</sup>

The curvilinear nature of  $R2$  is easy to explain. Both  $R2$  and  $R2^*$  depend upon the size and distribution of magnetic inhomogeneities.<sup>20,27</sup>  $R2$  should rise linearly with iron only if the size and cellular distribution of liver iron deposits are independent of HIC. In particular,  $R2$  becomes progressively less sensitive to magnetic fluctuations greater than the cellular scale because water molecules travel slowly across membrane boundaries. If severe iron loading produces proportionally greater magnetic inhomogeneities on the order of 100 micrometers or larger, then one would expect  $R2$  to "plateau" at high HIC.  $R2^*$  measurements are robust to long-range magnetic disturbances, thus one would expect a linear relationship between  $R2^*$  and iron over the entire physiologic range of iron deposition.<sup>27</sup>

There is much less published data on using liver  $R2^*$  to estimate liver iron. Anderson et al<sup>19</sup> found a negative logarithmic relationship between  $T2^*$  (the reciprocal of  $R2^*$ ) and biopsied liver iron

concentration. Translated to  $R2^*$  values, their data implied a near-linear rise of  $R2^*$  with HIC and slope double that observed in our study. However, the confidence intervals on their regression analysis were sufficiently broad that this slope difference was not statistically significant. Their study was limited by a minimum echo time of 2.2 milliseconds, compared with 0.8 milliseconds in our study. Inappropriately long echo times severely degrade estimates of liver  $T2$  or  $T2^*$  at high iron loads.<sup>16,31</sup> We believe that our improved agreement with liver biopsy and the high concordance between  $R2^*$  and nonlinear  $R2$  HIC estimates support our  $R2^*$  versus iron calibration.  $R2^*$  measurements also appear to have acceptable intermachine reproducibility,<sup>32,33</sup> although larger-scale validations will be necessary to determine whether  $R2^*$  measurements can be performed with the same machine independence recently demonstrated for  $R2$  measurements.<sup>18</sup>

Combined  $R2^*$  and  $R2$  HIC estimates were no more accurate than either alone, either by using  $R2'$  estimation or by simple averaging of the  $R2$  and  $R2^*$  HIC predictions. However, our study was relatively small and underpowered to detect changes less than 33% in confidence interval size. Whether or not it represents a statistically improved liver iron estimate, we find that  $R2$  and  $R2^*$  estimates tend to "bracket" the biopsied value, and having both estimates provides an additive degree of user confidence in the MRI prediction. Disparate  $R2$  and  $R2^*$  estimates prompt careful review of the patient's images for artifacts and flag the resultant value with a larger degree of uncertainty.

Despite the strong agreement of all the MRI methods with liver biopsy, the 95% confidence intervals for regression and Bland-Altman analyses appear large. The source of the error is at least 3-fold. (1) MRI measurements of  $R2$  and  $R2^*$  are imperfect. This effect is relatively small until iron concentrations exceed 30 mg/g. In general, MRI techniques have low interstudy variation (7.4%-8.3%). This value is comparable with iron measurement errors (COV, 7%) on reference iron standards (John Butz, Mayo Medical Laboratory; personal oral communication, January 2005). (2) Sensitivity of  $R2$  and  $R2^*$  to liver iron has some patient specificity because of interpatient variations in the size and susceptibility of iron deposits. (3) Liver biopsy is a relatively poor marker of "average" hepatic iron burden because of sampling variation. How large is this effect? Previous studies suggest a COV for liver biopsy ranging from 15% to 25% in healthy livers<sup>34,35</sup>; higher values have been described in diseased livers.<sup>36,37</sup> Our patient population was young and free from hepatitis C or significant fibrosis. Assuming a COV of only 15%, Bland-Altman confidence intervals for 2 biopsies from the same patient would be expected to be  $-41\%$  to  $41\%$ . These limits of agreement are similar to the results obtained by the MRI techniques in this paper (Table 1). Therefore, much of the disagreement between MRI and biopsy arises from the heterogeneity of iron deposition within the liver.

So is MRI a more *accurate* indicator of liver iron than biopsy? Probably not, at least not for patients with minimal liver disease. Even with *perfect* measurements of liver  $R2$  and  $R2^*$ , the calibration curve between  $R2$  and  $R2^*$  with iron is patient specific and may vary, subtly, with time or iron chelation. These patient-specific variations in the calibration curve are significantly larger than the measurement error. The magnitude of this error can be inferred by the limits of agreement between MRI estimates ( $-66\%$ - $43\%$ ); the limits of agreement are significantly larger than between MRI technique or liver biopsy alone ( $R2^*$ ,  $-46\%$ - $44\%$ ;  $R2$ ,  $-46\%$ - $34\%$ ). Comparable confidence intervals would be observed for a technique having a COV of

20%; hence, one must conclude that the sum of the patient-specific and measurement errors for MRI is of similar magnitude to the intrinsic variability of biopsy. This observation is in agreement with in vitro studies of R2 HIC estimation in which iron sampling error was eliminated.<sup>23</sup>

However even if MRI has only comparable accuracy to liver biopsy, it has many advantages. Interstudy variability is low, making it a good tool for serial evaluation of chelation efficacy. In our experience, patients are much more likely to agree to annual MRIs than annual biopsies, leading to closer monitoring. It is relatively inexpensive (~\$500) and can be performed at the same time as cardiac function and cardiac iron evaluation (T2\* or signal intensity ratio). From a pragmatic standpoint, management decisions do not rely on perfect determination of liver iron (Table 2). For this reason, liver biopsy for the sole purpose of iron determination has essentially disappeared from our institution. Biopsy is still

indicated when tissue histology is important for patient management. Furthermore, MRI HIC determination does not preclude liver biopsy if the HIC determination does not make clinical “sense” or is near an important therapeutic “boundary.”

Both R2 and R2\* MRI measurements using modified gradient and spin-echo imaging sequences produced highly accurate noninvasive estimates of hepatic iron over the entire clinically relevant range. HIC measurements by R2 and R2\* had equivalent accuracy, but combined measurements were not better than either one alone. We found a nonlinear R2-iron relationship consistent with the calibration curve observed by St Pierre et al,<sup>18</sup> demonstrating that MRI measurements of R2 and R2\* are robust and instrument-independent estimators of HIC. Using commonly available instrumentation, MRI measurement of R2 and R2\* provides a rapid, robust, and accurate method for estimation of hepatic iron concentration suitable for diagnosis and management of transfusional iron overload.

## References

- Angelucci E, Brittenham GM, McLaren CE, et al. Hepatic iron concentration and total body iron stores in thalassemia major [published erratum appears in N Engl J Med. 2000;343:1740]. N Engl J Med. 2000;343:327-331.
- Olivieri NF, Brittenham GM. Iron-chelating therapy and the treatment of thalassemia. Blood. 1997;89:739-761.
- Nielsen P, Engelhardt R, Duerken M, Janka GE, Fischer R. Using SQUID biomagnetic liver susceptometry in the treatment of thalassemia and other iron loading diseases. Transfus Sci. 2000;23:257-258.
- Fischer R, Longo F, Nielsen P, Engelhardt R, Hider RC, Piga A. Monitoring long-term efficacy of iron chelation therapy by deferiprone and desferrioxamine in patients with beta-thalassemia major: application of SQUID biomagnetic liver susceptometry. Br J Haematol. 2003;121:938-948.
- Fischer R, Tiemann CD, Engelhardt R, et al. Assessment of iron stores in children with transfusion siderosis by biomagnetic liver susceptometry. Am J Hematol. 1999;60:289-299.
- Brittenham GM, Farrell DE, Harris JW, et al. Magnetic-susceptibility measurement of human iron stores. N Engl J Med. 1982;307:1671-1675.
- Nielsen P, Fischer R, Engelhardt R, et al. Liver iron stores in patients with secondary haemosiderosis under iron chelation therapy with deferrioxamine or deferiprone. Br J Haematol. 1995;91:827-833.
- Stark DD, Bass NM, Moss AA, et al. Nuclear magnetic resonance imaging of experimentally induced liver disease. Radiology. 1983;148:743-751.
- Voskaridou E, Douskou M, Terpos E, et al. Magnetic resonance imaging in the evaluation of iron overload in patients with beta thalassemia and sickle cell disease. Br J Haematol. 2004;126:736-742.
- Engelhardt R, Langkowski JH, Fischer R, et al. Liver iron quantification: studies in aqueous iron solutions, iron overloaded rats, and patients with hereditary hemochromatosis. Magn Reson Imaging. 1994;12:999-1007.
- Wang ZJ, Haselgrove JC, Martin MB, et al. Evaluation of iron overload by single voxel MRS measurement of liver T2. J Magn Reson Imaging. 2002;15:395-400.
- Dixon RM, Styles P, al-Rafaie FN, et al. Assessment of hepatic iron overload in thalassemic patients by magnetic resonance spectroscopy. Hepatology. 1994;19:904-910.
- Kaltwasser JP, Gottschalk R, Schalk KP, Hartl W. Non-invasive quantitation of liver iron-overload by magnetic resonance imaging. Brit J Haematol. 1990;74:360-363.
- Mavrogeni SI, Gotsis ED, Markussis V, et al. T2 relaxation time study of iron overload in b-thalassemia. Magma. 1998;6:7-12.
- Papakonstantinou O, Kostaridou S, Maris T, et al. Quantification of liver iron overload by T2 quantitative magnetic resonance imaging in thalassemia: impact of chronic hepatitis C on measurements. J Pediatr Hematol Oncol. 1999;21:142-148.
- Brittenham GM, Badman DG. Noninvasive measurement of iron: report of an NIDDK workshop. Blood. 2003;101:15-19.
- Angelucci E, Giovagnoni A, Valeri G, et al. Limitations of magnetic resonance imaging in measurement of hepatic iron. Blood. 1997;90:4736-4742.
- 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.
- 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.
- Weisskoff RM, Zuo CS, Boxerman JL, Rosen BR. Microscopic susceptibility variation and transverse relaxation: theory and experiment. Magn Reson Med. 1994;31:601-610.
- Clark PR, Chua-anusorn W, St Pierre TG. Bi-exponential proton transverse relaxation rate (R2) image analysis using RF field intensity-weighted spin density projection: potential for R2 measurement of iron-loaded liver. Magn Reson Imaging. 2003;21:519-530.
- Clark PR, St Pierre TG. Quantitative mapping of transverse relaxivity (1/T(2)) in hepatic iron overload: a single spin-echo imaging methodology. Magn Reson Imaging. 2000;18:431-438.
- Clark PR, Chua-anusorn W, St Pierre TG. Proton transverse relaxation rate (R2) images of liver tissue: mapping local tissue iron concentrations with MRI. Magn Reson Med. 2003;49:572-575.
- Graham JM, Paley MN, Grunewald RA, Hoggard N, Griffiths PD. Brain iron deposition in Parkinson's disease imaged using the PRIME magnetic resonance sequence [published erratum appears in Brain. 2001;124:1258]. Brain. 2000;123:2423-2431.
- Miszkiel KA, Paley MN, Wilkinson ID, et al. The measurement of R2, R2\* and R2' in HIV-infected patients using the prime sequence as a measure of brain iron deposition. Magn Reson Imaging. 1997;15:1113-1119.
- 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.
- Tanimoto A, Oshio K, Suematsu M, Pouliquen D, Stark DD. Relaxation effects of clustered particles. J Magn Reson Imaging. 2001;14:72-77.
- Tanimoto A, Pouliquen D, Kreft BP, Stark DD. Effects of spatial distribution on proton relaxation enhancement by particulate iron oxide. J Magn Reson Imaging. 1994;4:653-657.
- Bulte JW, Miller GF, Vymazal J, Brooks RA, Frank JA. Hepatic hemosiderosis in non-human primates: quantification of liver iron using different field strengths. Magn Reson Med. 1997;37:530-536.
- Brooks RA, Moyn F, Gillis P. On T2-shortening by weakly magnetized particles: the chemical exchange model. Magn Reson Med. 2001;45:1014-1020.
- Ghugre N, Enriquez C, Coates TD, et al. Improved R2\* measurements in myocardial iron overload. J Cardiovasc Magn Reson. 2005;7:32-33.
- Westwood M, Anderson LJ, Firmin DN, et al. A single breath-hold multiecho T2\* cardiovascular magnetic resonance technique for diagnosis of myocardial iron overload. J Magn Reson Imaging. 2003;18:33-39.
- 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.
- Ambu R, Crisponi G, Sciort R, et al. Uneven hepatic iron and phosphorus distribution in beta-thalassemia. J Hepatol. 1995;23:544-549.
- 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.
- Villeneuve JP, Bilodeau M, Lepage R, Cote J, LeFebvre M. Variability in hepatic iron concentration measurement from needle-biopsy specimens. J Hepatol. 1996;25:172-177.
- Emond MJ, Bronner MP, Carlson TH, Lin M, Labbe RF, Kowdley KV. Quantitative study of the variability of hepatic iron concentrations. Clin Chem. 1999;45:340-346.