Blood First Edition Paper, prepublished online April 28, 2005; DOI 10.1182/blood-2004-10-3982

MRI R2 and R2* Mapping Accurately Estimates Hepatic Iron Concentration in Transfusion-Dependent Thalassemia and Sickle-Cell Disease Patients

John C Wood, MD, PhD[1,3], Cathleen Enriquez, BS[1], Nilesh Ghugre MS[3] , J. Michael Tyzka PhD[2], Susan Carson MSN[1], Marvin D Nelson, MD[3] and Thomas D Coates, MD[1]

[1]Department of Pediatrics, Childrens Hospital of Los Angeles, Los Angeles, CA, 90027-0034; [2]Department of Biology, California Institute of Technology, Pasadena, CA, 91125

[3]Department of Radiology, Childrens Hospital of Los Angeles, Los Angeles, CA, 90027-0081.

ABSTRACT

Measurements of hepatic iron content (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, non-invasive 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 manuscript compares 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. Importantly, the R2 calibration curve was similar to the curve generated by St. Pierre, et al, despite significant differences in technique and instrumentation. Combined R2 and R2* measurements did not yield more accurate results than either one alone. Both R2 and R2* can accurately measure hepatic iron concentration throughout the clinically relevant range of HIC with appropriate MRI acquisition techniques.

INTRODUCTION

Hepatic iron content (HIC) is used as a surrogate for total iron balance to guide chelation therapy in transfusion-dependent patients[1, 2]. 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 four such devices are currently operating worldwide [3-7]. The use of magnetic resonance imaging (MRI) relaxation time techniques to estimate liver iron content has been studied for nearly 20 years[8-15]. 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 hepatic iron content[16].

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[8-15], however these studies have been criticized for their small size, limited dynamic range, and inter-study calibration variability[16, 17]. In a study of over one hundred patients, St. Pierre et al found a curvilinear R2 relationship between R2 and biopsy HIC over the entire clinically relevant range[18]. More importantly, that study demonstrated measurement stability between multiple imaging platforms. R2 changes have also been qualitatively associated with cardiac iron deposition[9, 14].

There are fewer studies of R2* for iron quantitation. Anderson et al 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 [19]. 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[19-23]. Some investigators have even proposed that the difference between R2* and R2, called R2', make be a more specific marker of tissue iron[24, 25]. 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 normal controls. MRI was validated to biopsy-measured iron content 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).

METHODS

Patient Population: 102 patients underwent a total of 132 comprehensive iron evaluations between 8/2002 and 8/2004. Patients were primarily referred for cardiac T2* and cardiac function analysis but consented to MRI HIC assessment as well. 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 between biopsy and MRI examination was 4.3 ± 9.5, [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 ± 0.6 mg [5.1-6.3].

Liver iron concentrations in the patient population were quite high. We were unable to obtain HIC estimates from patients with non-transfusional iron overload, as was done in a previous study[26], 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 (9M,4F), ages 29.3 ± 12.3[12-50]. 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*age (units of μg/g dry wt, 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, i.e., $[Fe]^{\mu} = [Fe]^{95\%} / (1+1.96(COV))$. This yielded an age-matched mean iron estimate of 1.17 mg/g dry weight for our normal control population.

MRI techniques: MRI measurements were performed using a 4-element torso coil on a 1.5 T General Electric CVi scanner. Liver R2* was measured from a single mid-hepatic slice using a single echo, gradient echo sequence; TE was automatically stepped at .25 ms intervals from .8 to 4.8 ms in a single breath-hold. Other imaging parameters included a

field of view of 48 x 24, flip angle 20°, matrix 64 x 64, slice thickness 15 mm, bandwidth 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°-120° Hahn echo using TE s of 3.5, 5, 8, 12, 18 and 30 ms. A Hahn echo yielded significantly shorter echo times than a conventional 90°-180° 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 x 24, matrix 64 x 64, slice thickness 15 mm, gap 5 mm, and bandwidth 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:

$$S(TE) = A\, e^{-TE \cdot R2^*} + \mathbf{C} \quad (1)$$

The constant, **C**, was necessary to compensate for contributions from instrumentation noise and effects from iron-poor species such as blood and bile. Equation [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

$$[Fe]_{R2^*} = .0254 \times R2^* + 0.202 \quad (2)$$

$$[Fe]_{R2\text{-}L} = 0.148 \times R2 - 6.51 \quad (3)$$

$$[Fe]_{R2'} = 0.0329 \times R2' \quad (4)$$

where R2' is R2*-R2, $[Fe]_{R2*}$ is the HIC estimated from R2*, $[Fe]_{R2\text{-}L}$ is the HIC estimated from R2 using a linear fit, and $[Fe]_{R2'}$ is the HIC estimated from R2'.

To better characterize the nonlinear relationship of R2 versus iron, R2 was plotted against $[Fe]_{R2*}$ 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, given by

$$R2 = 6.88 + 26.06[Fe]^{0.701} - 0.438[Fe]^{1.402} \quad (5)$$

Estimates of liver iron by the St. Pierre R2 calibration, $[Fe]_{R2\text{-}SP}$, can be achieved from equation (5) by completing the square and algebraic manipulations to yields the following:

$$[Fe]_{R2\text{-}SP} = \left(29.75 - \sqrt{900.7 - 2.283R2}\right)^{1.424} \quad (6)$$

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 two measurements divided by the mean of two 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 to one another using a two-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 1-3 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, peri-portal fibrous expansion in 11 patients and minimal fibrosis in the remainder. Biopsy-measured HIC was uniformly distributed from 1.3-32.9 mg/g dry weight except for 1 patient with a 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 contents in the 102 iron overloaded patients (using average HIC calculated from equations 2 and 6) was 13.1 ± 12.3 [1.2-57.3 mg/g dry weight]. The patients who underwent liver biopsy tended to have higher MRI-HIC's 17.7 ± 12.1 [2.1-46.4 mg/g dry]. Though 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 between R2* and HIC was excellent from 1.3 to 32.9 mg/g dry weight. Regression analysis yielded a correlation coefficients of 0.97, slope of 37.4 Hz/(mg/g dry wt) and a y-intercept of 23.7 Hz. Normal controls had a mean R2* of 39.9 Hz ± 2.8 (SEM); this was plotted against their estimated HIC's (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 $[Fe]_{R2^*}$ and biopsy derived HIC's is shown in Table 1. There was no significant bias and confidence intervals were [-46%, 44%], comparable to a recently published R2 methodology[18].

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 demonstrates a slope of 6.54 Hz/ (mg/g dry wt), a y-intercept of 47.4 Hz, and a correlation coefficient of 0.98. Limits of agreement between $[Fe]_{R2\text{-}L}$ demonstrates statistically insignificant bias (-4%, p=0.48) and comparable 95% confidence intervals ([-55%, 46%], p=0.40) to the work of St Pierre et al [18]. Normal controls had a mean R2 of 38.2 Hz ± 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 normal controls (55.1 Hz compared with 38.2 Hz, p<0.0001).

Further evidence that the R2 calibration curve is nonlinear comes by plotting R2-iron relationship for all 132 iron examinations. Figure 3 demonstrates R2 as a function of HIC, measured by biopsy (+ signs) and by $[Fe]_{R2*}$ (solid dots). With the additional examinations, the nonlinear trend is quite obvious. Curvature is only pronounced for HIC's < 7 mg/g but the trend passes through the normal control estimates. The bold line represents the calibration curve proposed by St Pierre et al (equation 5). This curve fits the low iron behavior well, with a correlation coefficients of 0.97 with respect to biopsy HIC and 0.96 with respect to HIC by R2*. The limits of agreement between $[Fe]_{R2\text{-}SP}$ and $[Fe]_{biopsy}$ were smaller than for a linear calibration, [-46%, 34%] vs [-55%, 46%], although this difference was not statistically significant (p=0.15). 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 normal controls.

Agreement between $[Fe]_{R2*}$ (equation 2) and $[Fe]_{R2\text{-}SP}$ (equation 6) is demonstrated by scattergram in Figure 4, along with its regression line. Both $[Fe]_{R2*}$ and $[Fe]_{R2\text{-}SP}$ estimates rise at the same rate (slope = 1.01 ± 0.02) and the correlation between them is 0.94. Despite this, $[Fe]_{R2\text{-}SP}$ exhibited a -11% bias toward $[Fe]_{R2*}$ and confidence intervals were broad [-66%, 43%]. The bias is greatest between HIC's 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 HIC's greater than 30 mg/g.

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 < 7 mg/g, corresponding to the greatest nonlinearity in R2 measurements. Bland-Altman agreement of $[Fe]_{R2'}$ (equation 3) with

biopsy values is comparable to 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 Oliveri[2]: 1) HIC <3.2 mg/g, concerns for chelator toxicity, 2) 3.2-7.0 mg/g, optimal chelation range, 3) 7-15 mg/g, elevated hepatic iron levels, and 4) HIC > 15 mg/g, markedly increased iron levels and potential cardiotoxicity. Bold lines indicate the division between these classifications. For this population, which was somewhat skewed toward heavy iron overload, there was only one classification "error" with a patient having a 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 was representative of our patient population having MRI-estimated iron concentration of 15.9 ± 13.9 [1.5-41 mg/g dry weight].

DISCUSSION

R2 and R2* methods have theoretical advantages and disadvantages to 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, 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, i.e., they more accurately reflect bulk magnetic susceptibility of tissues [20, 27, 28]. R2* measurements can also be performed in a single breath-hold while R2 methods take 5-20 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%, 34%] and $[Fe]_{R2*}$ of [-46%, 44%] compare favorably with the limits of agreement of [56%, 50%] found by St. Pierre et al 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, including younger patients, absence of hepatitis C or significant liver fibrosis, the use of an entire, fresh liver core for assessment, and the examinations were performed at a single center. We also had proportionally greater patients with liver irons > 10 mg/g where our MRI techniques are quite accurate. In fact, similarities between the two 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's 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 suggests a slightly flatter

calibration curvature for HIC's 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 1-50 mg/g, the curvilinearity is obvious. Other factors could also play a role in the observed inter-study variation. Some were performed at different magnetic field strengths; this has a profound effect on the magnitude and shape of the calibration curve [29]. R2 measurements using a train of echoes (also known as CPMG sequences) will be lower than those performed by single spin-echo techniques and will vary with echo-spacing [30]. Lastly choice of fitting algorithm can impact the estimated R2 or R2* [31].

The curvilinear nature of R2 is easy to explain. Both R2 and R2* depend upon the size and distribution of magnetic inhomogeneities [20, 27]. R2 should only rise linearly with iron 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 a hundred microns 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 [27].

There is much less published data on using liver R2* to estimate liver iron [19]. Anderson et al 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 m/s, compared with 0.8 ms in our study. Inappropriately long echo times severely degrade estimates of liver T2 or T2* at high iron loads [16, 31]. 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[32, 33] 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 [18].

Combined R2* and R2 HIC estimates were no more accurate than either one 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 < 33% in confidence interval size. Whether it represents a statistically improved liver iron estimate or not, 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 measurement of R2 and R2* are imperfect. This effect is relatively small until iron concentrations exceed 30 mg/g. In general, MRI techniques have low inter-study variation (7.4%-8.3%). This value is comparable to iron measurement errors (COV 7%) on reference iron standards (John Butz, Mayo Medical Laboratory). #2) Sensitivity of R2 and R2* to liver iron has some patient-specificity because of inter-patient 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 coefficient of variation (COV) for liver biopsy ranging from 15-25% in healthy livers [34, 35]; higher values have been described in diseased livers [36, 37]. Our patient population was young and free from hepatitis C or significant fibrosis. Assuming a COV of only 15% Bland-Altman confidence intervals for two biopsies from the same patient would be expected to be [–41%, 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%]), which is significantly larger than between either MRI technique and liver biopsy alone (R2* [-46%, 44%] and 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 are of similar magnitude to the intrinsic variability of biopsy. This observation is in agreement with in-vitro studies of R2 HIC estimation where iron sampling error was eliminated [23].

However even if MRI only has comparable accuracy to liver biopsy, it has many advantages. Inter-study 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 MRI's 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 don't rely on perfect determination of liver iron (see 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".

CONCLUSION

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., demonstrating that MRI measurement 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.

ACKNOWLEDGEMENTS

This work supported by the National Heart Lung and Blood Institute (1R01 HL75592-01A1) and National Center for Research Resources (General Clinical Research Center (GCRC) RR00043-43) at the National Institute of Health as well as Novartis Pharma, AG, the Whitaker Foundation, and the Department of Pediatrics at Children's Hospital of Los Angeles.

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

Method	Bias	Standard deviation	p-value vs St Pierre	95% Confidence
R2*	1%	23%	0.21	[-46%, 44%]
R2 linear	-4%	26%	0.43	[-55%, 46%]
R2 nonlinear	-6%	20%*	0.08	[-46%, 34%]
R2'	-8%	23%	0.24	[-54%, 38%]
Avg(R2,R2*)	-3%	20%	0.08	[-43%, 37%]
R2 St. Pierre	-3%	27%	-----	[-56%, 50%]

* p=0.15 versus R2 linear

Table 2

Hepatic Iron Concentration

Patient	Biopsy	R2*	R2	R2'	R2-R2*	
RW	1.3	2.2	2.0	1.0	2.1	*"Low"*
HE	4.4	6.1	3.5	5.5	5.5	
CS	4.6	5.3	4.7	4.1	5.1	
GZ	5.9	4.5	4.9	3.1	4.6	*"Optimal"*
FM	6.0	3.8	3.5	6.4	3.6	
VM	7.8	10.9	5.9	11.2	8.4	
LR	8.3	7.6	8.1	7.6	7.9	
VL	12.7	9.3	8.4	8.5	8.9	*"Increased"*
BB	13.4	12.1	12.7	11.2	12.4	
MF	14.8	12.2	13.0	10.5	12.6	
LQ	16.6	19.0	18.9	19.1	19.0	
YP	19.2	18.1	17.6	17.8	17.9	
BJ2	21.3	20.4	20.2	20.7	22.7	
DT	21.9	21.6	20.6	21.8	21.1	
BS	23.1	21.4	22.2	21.1	21.8	
BJ1	25.5	24.0	25.0	24.0	24.5	*"High"*
NA2	27.3	32.5	29.4	35.8	31.0	
TS	29.0	24.2	28.1	23.4	26.2	
NA1	29.6	36.1	26.3	36.0	31.2	
BL	30.0	29.0	34.2	28.8	31.6	
CL	32.9	33.8	33.5	34.9	33.7	
ES	57.8	46.0	38.2	50.3	42.3	
MB	16.2	15.9	-----	-----	-----	

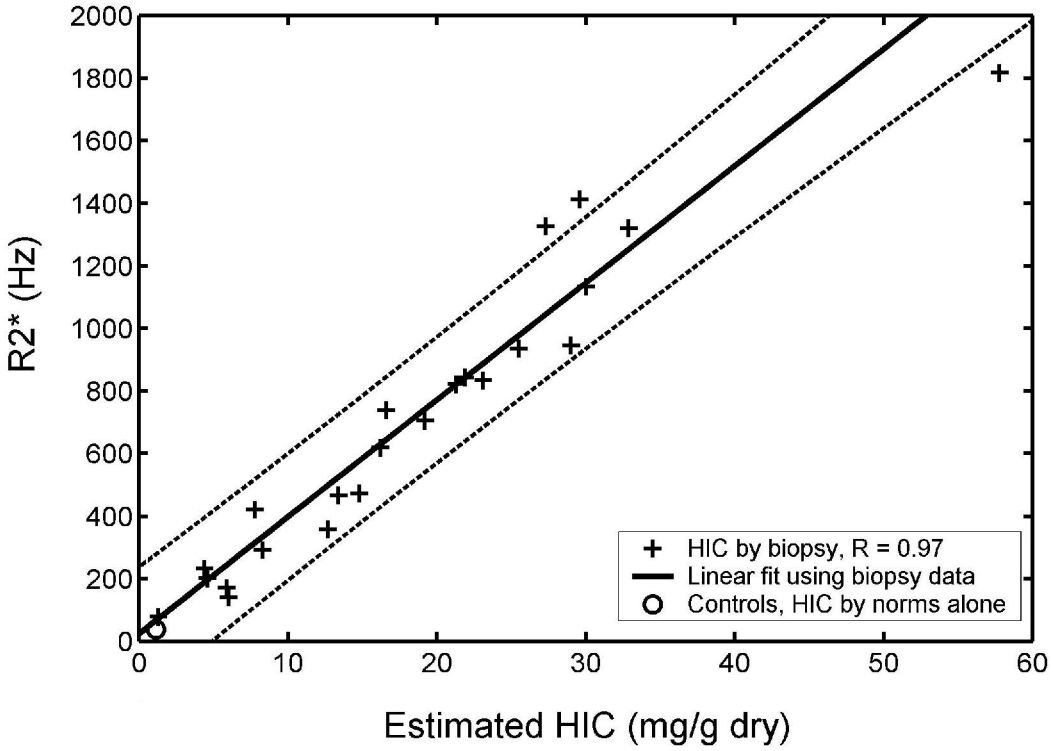
2000
1800
1600
1400
1200
1000
800
600
400
200
0
R2* (Hz)
0
10
20
30
40
50
60
Estimated HIC (mg/g dry)
HIC by biopsy, R = 0.97
Linear fit using biopsy data
Controls, HIC by norms alone

Figure 1

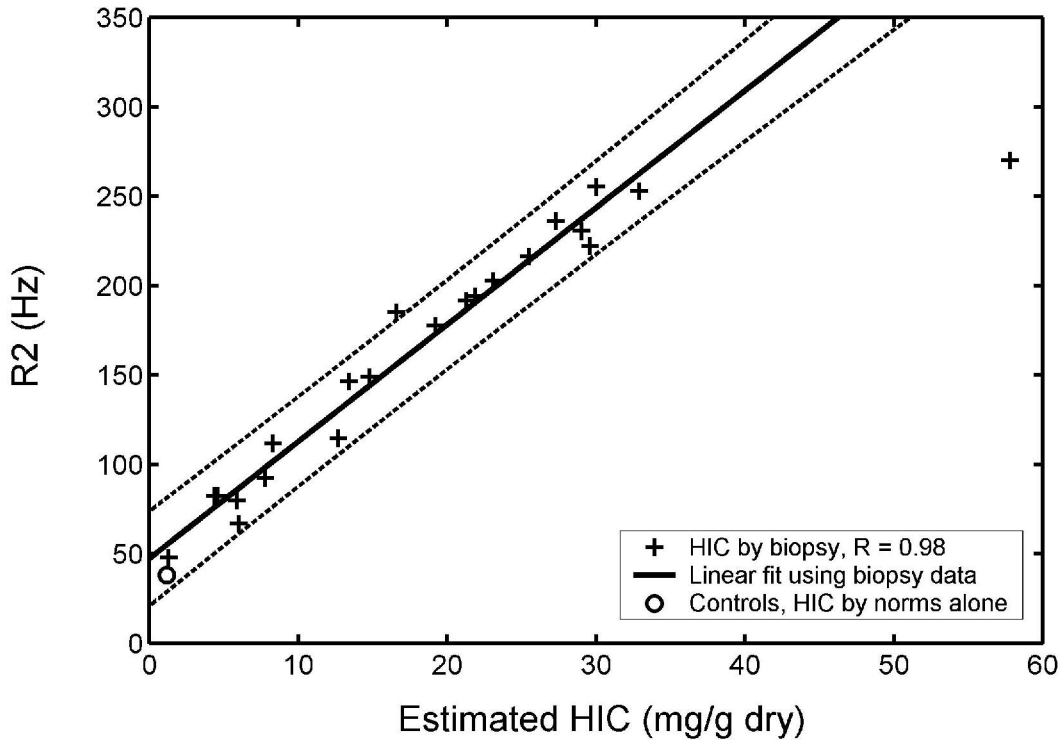
R2 (Hz)
Estimated HIC (mg/g dry)
HIC by biopsy, R = 0.98
Linear fit using biopsy data
Controls, HIC by norms alone
0
50
100
150
200
250
300
350
0
10
20
30
40
50
60

Figure 2

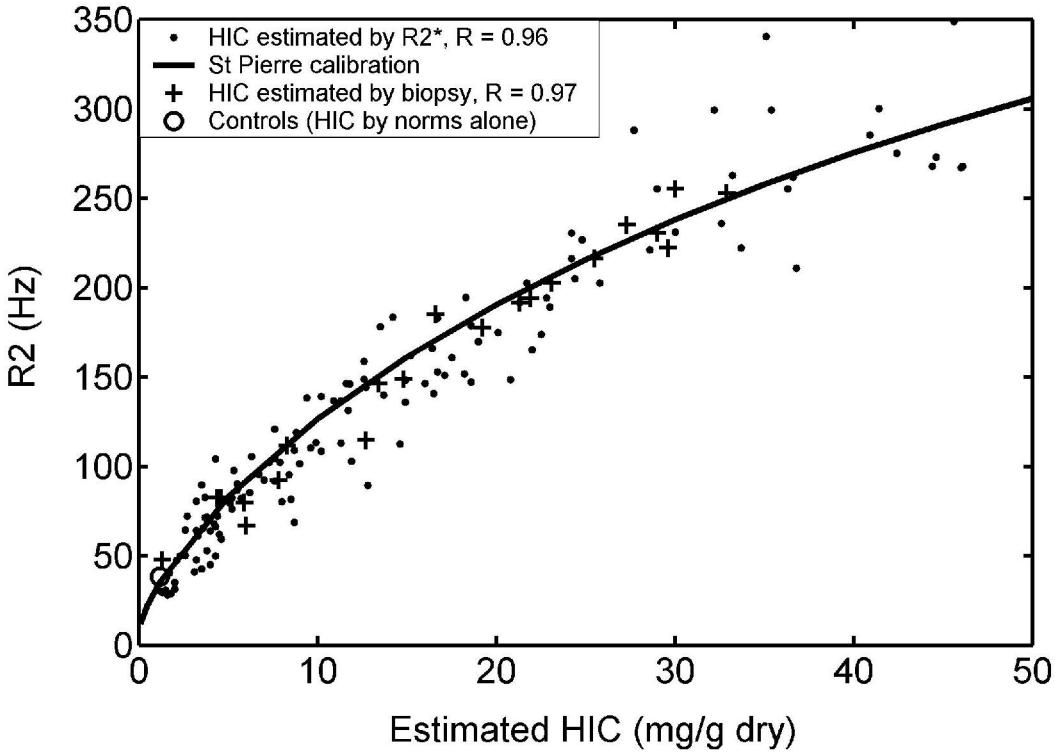
HIC estimated by R2*, R = 0.96
St Pierre calibration
HIC estimated by biopsy, R = 0.97
Controls (HIC by norms alone)
R2 (Hz)
Estimated HIC (mg/g dry)
0
50
100
150
200
250
300
350
0
10
20
30
40
50

Figure 3

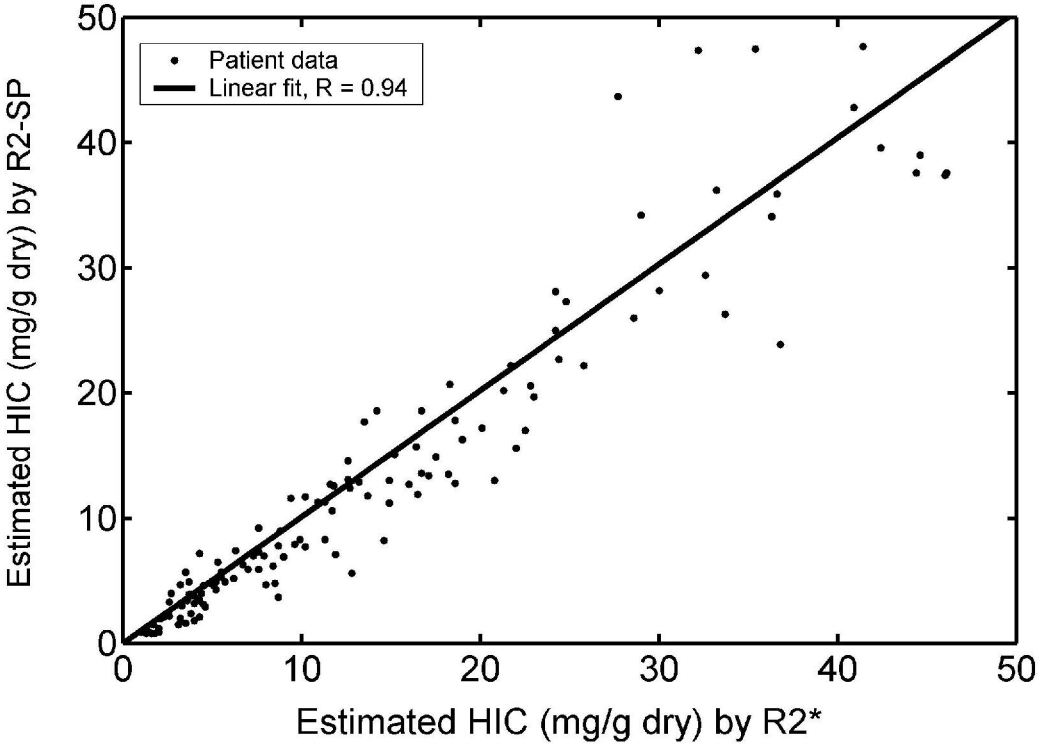
Patient data
Linear fit, R = 0.94
Estimated HIC (mg/g dry) by R2-SP
Estimated HIC (mg/g dry) by R2*
0
10
20
30
40
50

Figure 4

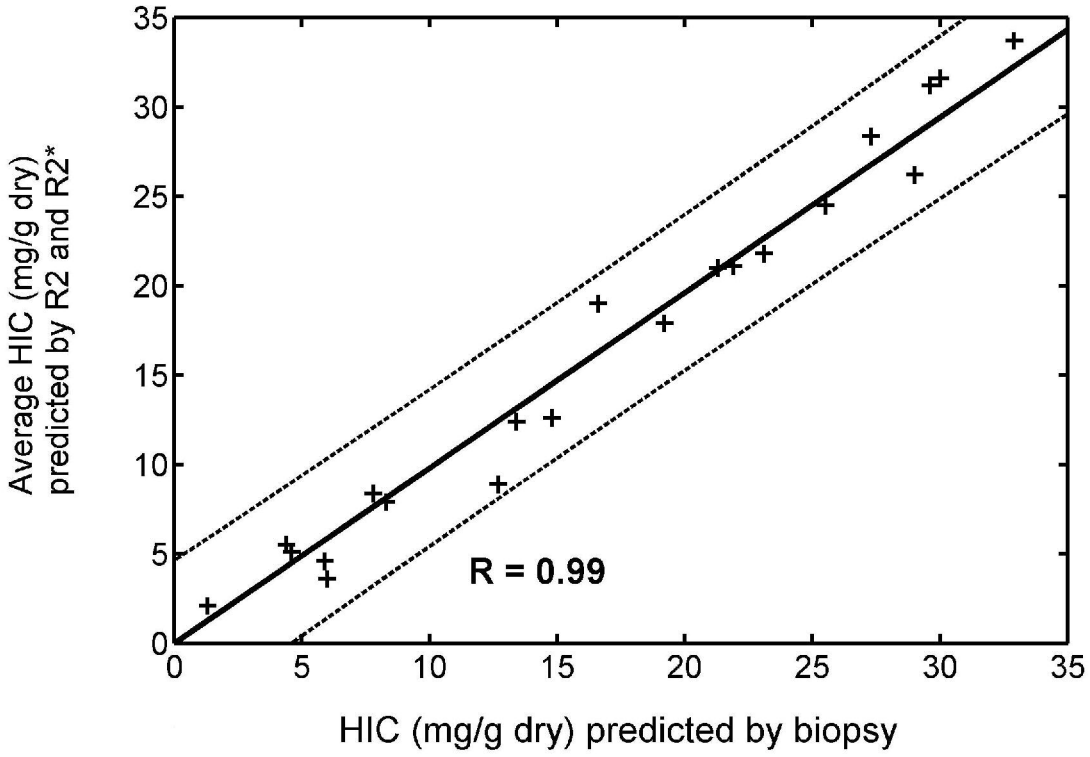
Average HIC (mg/g dry) predicted by R2 and R2*
HIC (mg/g dry) predicted by biopsy
R = 0.99
0
5
10
15
20
25
30
35

Figure 5

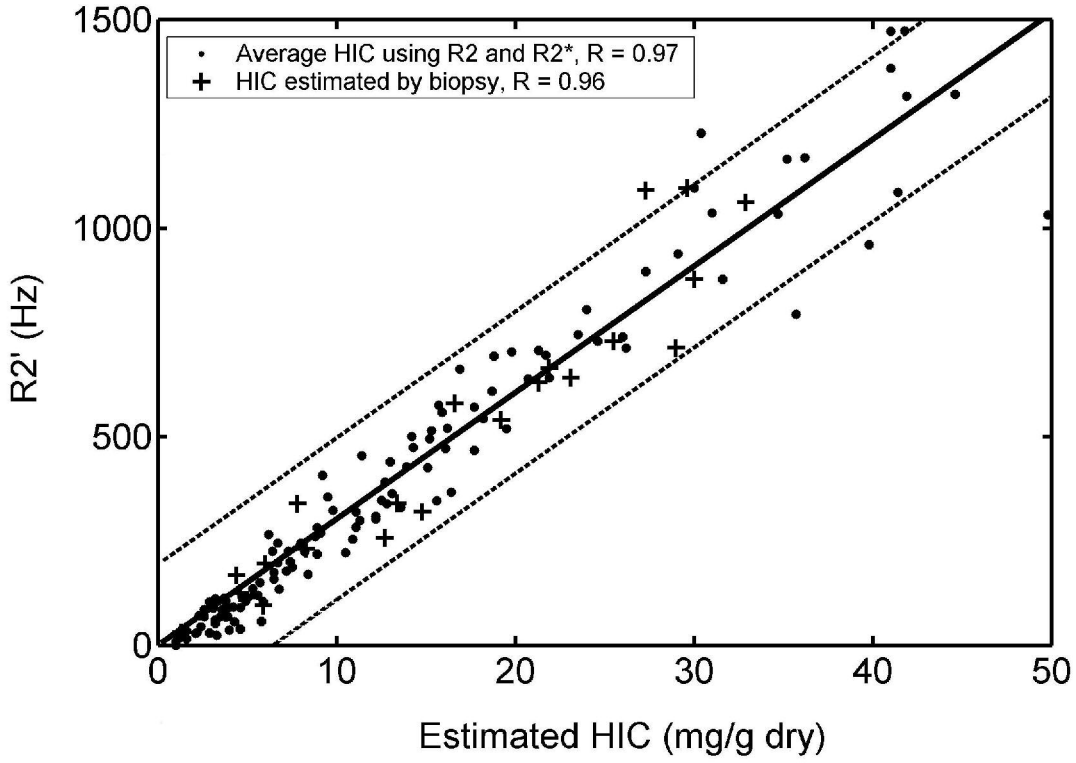
Average HIC using R2 and R2*, R = 0.97
HIC estimated by biopsy, R = 0.96
R2' (Hz)
Estimated HIC (mg/g dry)
0
500
1000
1500
0
10
20
30
40
50

Figure 6

FIGURE LEGENDS

1. A. Plot of transverse relaxivity R2* (1/T2*) versus biopsied hepatic iron concentration(HIC) in 21 patients(23 biopsies). R2* has units of Hz and HIC has units of mg/g dry weight of liver. R-value was 0.97 and dashed lines indicate 95% prediction intervals for the regression. Average R2* value for 13 normal controls are shown for comparison (open circle), plotted using a HIC value estimated from normative data (no biopsy). Repeat MRI and biopsy examinations as well as control data were excluded from statistical calculations.
2. A. Plot of transverse relaxivity R2 (1/T2) versus biopsy-measured hepatic iron concentration (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 normal controls is shown by the open circle, plotted using a HIC value estimated from normative data (no biopsy). Repeat MRI and biopsy examinations as well as control data were excluded from statistical calculations.
3. 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 a 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[18]. 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.

4. Comparison of iron content estimated by R2 (St. Pierre calibration, equation 6) and by R2* (equation 2). Regression slope is 1.01 ± 0.02 with a correlation coefficient of 0.94. Despite this, the HIC by R2 has a 11% bias relative to values predicted by R2* and limits of agreement are broader than for corresponding comparison with biopsy.
5. Average HIC predicted by R2 (equations 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). Regression slope is 0.98 and R-value is 0.99. However, limits of agreement are not improved from the individual measurements (Table 1).
6. Plot of R2' versus HIC estimated by MRI (average of equations 2 and 6) and by biopsy. 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 to HIC measurements by R2* alone (Table 1).

REFERENCES

1. Angelucci, E., et al., *Hepatic iron concentration and total body iron stores in thalassemia major.* N Engl J Med, 2000. **343**(5): p. 327-31.
2. Olivieri, N.F. and G.M. Brittenham, *Iron-chelating therapy and the treatment of thalassemia.* Blood, 1997. **89**(3): p. 739-61.
3. Nielsen, P., et al., *Using SQUID biomagnetic liver susceptometry in the treatment of thalassemia and other iron loading diseases.* Transfus Sci, 2000. **23**(3): p. 257-8.
4. Fischer, R., et al., *Monitoring long-term efficacy of iron chelation therapy by deferiprone and desferrioxamine in patients with beta-thalassaemia major: application of SQUID biomagnetic liver susceptometry.* Br J Haematol, 2003. **121**(6): p. 938-48.
5. Fischer, R., et al., *Assessment of iron stores in children with transfusion siderosis by biomagnetic liver susceptometry.* Am J Hematol, 1999. **60**(4): p. 289-99.
6. Brittenham, G.M., et al., *Magnetic-susceptibility measurement of human iron stores.* N Engl J Med, 1982. **307**(27): p. 1671-5.
7. Nielsen, P., et al., *Liver iron stores in patients with secondary haemosiderosis under iron chelation therapy with deferoxamine or deferiprone.* Br J Haematol, 1995. **91**(4): p. 827-33.
8. Stark, D.D., et al., *Nuclear magnetic resonance imaging of experimentally induced liver disease.* Radiology, 1983. **148**(3): p. 743-51.

9. Voskaridou, E., et al., *Magnetic resonance imaging in the evaluation of iron overload in patients with beta thalassaemia and sickle cell disease.* Br J Haematol, 2004. **126**(5): p. 736-42.

10. Engelhardt, R., et al., *Liver iron quantification: studies in aqueous iron solutions, iron overloaded rats, and patients with hereditary hemochromatosis.* Magn Reson Imaging, 1994. **12**(7): p. 999-1007.

11. Wang, Z.J., et al., *Evaluation of iron overload by single voxel MRS measurement of liver T2.* J Magn Reson Imaging, 2002. **15**(4): p. 395-400.

12. Dixon, R.M., et al., *Assessment of hepatic iron overload in thalassemic patients by magnetic resonance spectroscopy.* Hepatology, 1994. **19**: p. 904-910.

13. Kaltwasser, J.P., et al., *Non-invasive quantitation of liver iron-overload by magnetic resonance imaging.* Brit. J. Haem., 1990. **74**: p. 360-363.

14. Mavrogeni, S.I., et al., *T2 relaxation time study of iron overload in b-thalassemia.* Magma, 1998. **6**(1): p. 7-12.

15. Papakonstantinou, O., 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**(2): p. 142-8.

16. Brittenham, G.M. and D.G. Badman, *Noninvasive measurement of iron: report of an NIDDK workshop.* Blood, 2003. **101**(1): p. 15-9.

17. Angelucci, E., et al., *Limitations of magnetic resonance imaging in measurement of hepatic iron.* Blood, 1997. **90**(12): p. 4736-42.

18. St Pierre, T.G., et al., *Non-Invasive Measurement and Imaging of Liver Iron Concentrations Using Proton Magnetic Resonance.* Blood, 2004.

19. Anderson, L.J., et al., *Cardiovascular T2-star (T2*) magnetic resonance for the early diagnosis of myocardial iron overload.* Eur Heart J, 2001. **22**(23): p. 2171-9.

20. Weisskoff, R.M., et al., *Microscopic susceptibility variation and transverse relaxation: theory and experiment.* Magn Reson Med, 1994. **31**(6): p. 601-10.

21. Clark, P.R., W. Chua-anusorn, and T.G. St Pierre, *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**(5): p. 519-30.

22. Clark, P.R. and T.G. St Pierre, *Quantitative mapping of transverse relaxivity (1/T(2)) in hepatic iron overload: a single spin-echo imaging methodology.* Magn Reson Imaging, 2000. **18**(4): p. 431-8.

23. Clark, P.R., W. Chua-Anusorn, and T.G. St Pierre, *Proton transverse relaxation rate (R2) images of iron-loaded liver tissuepping local tissue iron concentrations with MRI.* Magn Reson Med, 2003. **49**(3): p. 572-5.

24. Graham, J.M., et al., *Brain iron deposition in Parkinson's disease imaged using the PRIME magnetic resonance sequence.* Brain, 2000. **123 Pt 12**: p. 2423-31.

25. Miszkiel, K.A., 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**(10): p. 1113-9.

26. St Pierre, T.G., et al., *Noninvasive measurement and imaging of liver iron concentrations using proton magnetic resonance.* Blood, 2005. **105**(2): p. 855-61.

27. Tanimoto, A., et al., *Relaxation effects of clustered particles.* J Magn Reson Imaging, 2001. **14**(1): p. 72-7.

28. Tanimoto, A., et al., *Effects of spatial distribution on proton relaxation enhancement by particulate iron oxide.* J Magn Reson Imaging, 1994. **4**(5): p. 653-7.

29. Bulte, J.W., et al., *Hepatic hemosiderosis in non-human primates: quantification of liver iron using different field strengths.* Magn Reson Med, 1997. **37**(4): p. 530-6.

30. Brooks, R.A., F. Moiny, and P. Gillis, *On T2-shortening by weakly magnetized particles: the chemical exchange model.* Magn Reson Med, 2001. **45**(6): p. 1014-20.

31. Ghugre, N., et al., *Improved R2* Measurements in Myocardial Iron Overload.* Journal of Cardiovascular Magnetic Resonance, 2005. **7**(1): p. 32-33.

32. Westwood, M., et al., *A single breath-hold multiecho T2* cardiovascular magnetic resonance technique for diagnosis of myocardial iron overload.* J Magn Reson Imaging, 2003. **18**(1): p. 33-9.

33. Westwood, M.A., et al., *Interscanner reproducibility of cardiovascular magnetic resonance T2* measurements of tissue iron in thalassemia.* J Magn Reson Imaging, 2003. **18**(5): p. 616-20.

34. Ambu, R., et al., *Uneven hepatic iron and phosphorus distribution in beta-thalassemia.* J Hepatol, 1995. **23**(5): p. 544-9.

35. Crisponi, G., et al., *Does iron concentration in a liver needle biopsy accurately reflect hepatic iron burden in beta-thalassemia?* Clin Chem, 2000. **46**(8 Pt 1): p. 1185-8.

36. Villeneuve, J.P., et al., *Variability in hepatic iron concentration measurement from needle-biopsy specimens.* J Hepatol, 1996. **25**(2): p. 172-7.

37. Emond, M.J., et al., *Quantitative study of the variability of hepatic iron concentrations.* Clin Chem, 1999. **45**(3): p. 340-6.