
Timothy G. St. Pierre¹, Paul R. Clark¹, Wanida Chua-anusorn¹, Adam J. Fleming¹, Gary P. Jeffrey²³, John K. Olynnyk², Pensri Pootrakul⁴, Erin Robins⁵, Robert Lindeman⁶

¹School of Physics and ²School of Medicine and Pharmacology, The University of Western Australia, Perth, AUSTRALIA,
³Department of Gastroenterology and Hepatology, Sir Charles Gairdner Hospital, Perth, AUSTRALIA
⁴Thalassemia Research Center, Mahidol University, Bangkok, THAILAND
⁵SKG Radiology, St. John of God Hospital, Perth, AUSTRALIA
⁶Dept of Haematology, Prince of Wales Hospital, Sydney, AUSTRALIA

Abbreviated title: Non-invasive measurement of liver iron concentrations

Word counts: total text 4817 abstract 199

This research was partly funded by the National Health and Medical Research Council (Project Grant 211947), Wellcome Trust (Collaborative Research Initiative Grant 068613), and the National Research Council of Thailand.

Section: Transfusion medicine

Several of the authors (TS, PC, WC) have declared a financial interest in a company whose potential product was studied in the present work.

Two of the authors (PC, WC) are employed by a company whose potential product was studied in the present work.

Correspondence:

A/Prof. Timothy G. St. Pierre
School of Physics, M013
The University of Western Australia
Crawley
Western Australia 6009
Australia

Tel: +(618) 9380 2747
Fax: +(618) 9380 1014
E-mail: stpierre@physics.uwa.edu.au
ABSTRACT

Measurement of liver iron concentrations (LIC) is necessary for a range of iron loading disorders such as hereditary hemochromatosis, thalassemia, sickle cell disease, aplastic anemia, and myelodysplasia. Currently, chemical analysis of needle biopsy specimens is the most common accepted method of measurement. This study presents a readily available non-invasive method of measuring and imaging LIC in vivo using clinical 1.5 T magnetic resonance imaging units. Mean liver proton transverse relaxation rates (R\textsubscript{2}) were measured for 105 human subjects. A value for the LIC for each subject was obtained by chemical assay of a needle biopsy specimen. High degrees of sensitivity and specificity of R\textsubscript{2} to biopsy LIC were found at the clinically significant LIC thresholds of 1.8, 3.2, 7.0 and 15.0 mg Fe/g dry tissue. A calibration curve relating liver R\textsubscript{2} to LIC has been deduced from the data covering the range of LIC from 0.3 to 42.7 mg Fe/g dry tissue. Proton transverse relaxation rates in aqueous paramagnetic solutions were also measured on each magnetic resonance imaging unit to ensure instrument-independent results. Measurements of proton transverse relaxivity of aqueous MnCl\textsubscript{2} phantoms on 13 different magnetic resonance imaging units using the method yielded a coefficient of variation of 2.1%.

Key Words: magnetic resonance imaging; iron; liver; hemochromatosis; thalassemia
INTRODUCTION

Body iron loading is associated with disorders such as hereditary hemochromatosis, thalassemia, sickle cell disease, aplastic anemia, myelodysplasia and others. The body iron burden is a principal determinant of clinical outcome in all forms of systemic iron overload, whether from red cell transfusion, from increased dietary iron absorption, or both. Accurate assessment of the body iron burden is essential for managing iron-chelation therapy to prevent iron toxicity while avoiding the adverse effects of excess chelator administration. In hereditary and non-hereditary forms of hemochromatosis, determination of the magnitude of body iron stores permits identification of individuals at risk from iron induced organ damage who would benefit from phlebotomy therapy.

The simplest methods available for assessment of body iron levels are biochemical measurements of the serum iron concentration, transferrin saturation, and ferritin concentration. However, serum biochemical tests can be confounded by factors such as infection, inflammation, and malignancy and do not accurately reflect tissue iron levels. The reference method for evaluating the magnitude of body iron load in systemic iron overload is measurement of the liver iron concentration (LIC). The most direct clinical method of measuring LIC is through chemical analysis of needle biopsy specimens. The biopsy sample can also be used for detection of liver fibrosis and cirrhosis, which have important prognostic implications for survival and risk of hepatocellular carcinoma. However, the measurement of LIC and detection of fibrosis or cirrhosis in biopsy specimens are subject to sampling variability, owing mainly to the
small size of the biopsy relative to the whole liver \(^6\)\(^-\)\(^8\). The variation in LIC throughout the liver increases as iron loading increases and with the development of cirrhosis \(^7\). Coefficients of variation (CV) for multiple needle biopsy measurements of LIC from individual livers range from an average of 19\% for disease-free liver \(^7\)\(^,\)\(^8\) to an average greater than 40\% for end-stage liver disease \(^7\)\(^,\)\(^8\) for typical needle biopsy sample dry masses of less than 4 mg. The analytical component of the variability has been estimated to be in the region of 3 to 7\% \(^7\)\(^,\)\(^9\). Furthermore, the invasive nature and risks associated with liver needle biopsy preclude serial observations.

Here we report a new non-invasive method for the measurement and imaging of liver iron concentrations \textit{in vivo} through the measurement of tissue proton transverse relaxation rates (R\(_2\)) using clinical magnetic resonance imaging instruments.

**METHODS**

**Magnetic resonance imaging**

\textit{Single-spin-echo image acquisition:} Magnetic resonance imaging on human subjects was conducted on five 1.5 T whole body imaging units (Siemens MAGNETOM Vision Plus (n = 4) and Siemens SONATA (n = 1)). Phased array torso coils were used for signal detection. Axial images were acquired with a multi-slice single-spin-echo (SSE) pulse sequence, with a pulse repetition time TR of 2500 ms, spin echo times TE of 6, 9, 12, 15, and 18 ms, and slice thickness = 5 mm. A matrix size of 256 was used with typical fields
of view being between 350 and 400 mm (exact dimensions depending on subject size). Each spin-echo sequence was run with fixed gain settings determined by the TE = 6 ms acquisition. Data were acquired in half Fourier mode in order to reduce measurement time with one acquisition. No fat suppression was used. A 1000 mL bag of Hartmann’s solution (compound sodium lactate, Viaflex®) was imaged with both the phantoms and human subjects to provide an external long T$_2$ reference for the correction of instrumental gain drift.

For liver studies, each subject was positioned so that the liver was located central to the phased array torso coil. Slices (n = 19) were collected for each subject, with the gap between slices adjusted to enable entire coverage of the liver (minimum gap size 5 mm).

Proton transverse relaxation rate (R$_2$) imaging: R$_2$ values were calculated throughout a liver slice by curve fitting the equation for the bi-exponential decay in transverse magnetisation following a SSE pulse sequence to the voxel intensity data as a function of TE$^{10}$. A mean R$_2$ value was calculated for each voxel by summation of the fast and slow components of the proton transverse relaxation rate weighted by their relative population densities as described elsewhere$^{10}$. To reduce image noise, the voxel intensities were smoothed over a 5x5 window kernel prior to curve fitting. Respiratory ghosting in the SSE images was reduced prior to generation of the R$_2$ images using methods described elsewhere$^{11}$. The generation of the liver R$_2$ images is described in greater detail elsewhere$^{12}$.

Selection of region of interest for R$_2$ analysis: For each subject, the largest axial slice of the liver was selected for R$_2$ image analysis. Given the heterogeneity of LIC and R$_2$
within the human liver, the \( R_2 \) measurement and needle biopsy site were approximately co-located to determine the relationship between \( R_2 \) and LIC. A lateral region of the right lobe of the liver bounded by its surface and a sagittal plane 35 mm medial to its most lateral surface point was used to calculate a mean \( R_2 \) value (\(<R_2>\)) for purposes of generating a calibration curve relating \( R_2 \) to LIC. To quantify the heterogeneity in \( R_2 \) for a subject, the entire slice of the liver was used for calculation of the standard deviation of \( R_2 \) (\( \sigma_{R_2} \)).

**Phantoms**

The precision and accuracy of \( R_2 \) measurements made using each MR scanner were assessed by measuring a series of manganese chloride (MnCl\(_2\)) solutions with different concentrations, prepared by serial dilution of a stock solution with distilled water. Concentrations ranged from 0.1 mM to 3.2 mM, which provided \( R_2 \) values in the range encompassed by healthy through to highly iron loaded liver. The phantoms were also measured on a variety of other makes and models of 1.5 T MRI scanners to assess the variation in the results of \( R_2 \) measurement methodology on different types of scanner. The instruments on which the \( R_2 \) measurements of the phantoms were tested comprised Siemens Magnetom Vision (n=5), Siemens Symphony (n=2), Siemens Sonata (n=1), Philips Intera (n=3), and General Electric Signa (n=2).

**Human Subjects**

All subjects gave written consent to participation in the study. All procedures on subjects were approved by the Human Ethics Committees at The University of Western Australia.
and St John of God Hospital and also by The Committee on Human Rights Related to Human Experimentation at Mahidol University, Bangkok, Thailand.

*R2-LIC calibration:* Subjects included patients who were about to undergo liver needle biopsy in the assessment of iron overload disorder or liver disease by their clinicians. The liver biopsy was used for routine histological examination and LIC measurement. MRI scanning was scheduled as close as possible to the liver biopsy procedure (a few days) or within 1 to 2 months for those volunteers who did not warrant clinical treatment for iron overload.

For the iron-loaded subjects, two major conditions were included in the study: hereditary hemochromatosis and thalassemia disorders. The hereditary hemochromatosis (HH) group consisted of subjects homozygous for the C282Y mutation on the HFE gene (n = 23, age range 17-74y). The thalassemia group consisted of β-thalassemic subjects who had been treated with regular blood transfusion and chelation therapy (n = 9; age range 8-36y) and β-thalassemic/Hb E subjects who had not received regular blood transfusion nor chelation therapy (n = 41, age range 12-63 y). The non-iron-loaded group consisted of subjects with hepatitis. From the 32 volunteers in this group, there were 29 with hepatitis C. Other cases of hepatitis were alcohol induced (n = 2) and drug induced (n = 1). Three of the hepatitis C subjects were heterozygous for the C282Y mutation of the HFE gene.

*Reproducibility tests (precision) of liver R2 measurements:* Subjects included 3 healthy volunteers, 5 with β-thalassemia major and 2 with hereditary hemochromatosis. Each
volunteer was measured on two MR scanners (both Siemens Magnetom Vision), the two measurements being made on consecutive days. The entire cross-section of the largest liver slice was used for the determination of the mean R₂ value in each case.

**Determination of liver biopsy iron concentration**

The chemical analysis for LIC measurement was conducted with atomic absorption spectrometry after acid digestion (4 laboratories). All samples had dry weights greater than 0.4 mg. Quality control studies for inter-laboratory assay were first performed using standard reference liver material (National Bureau of Standards BL1577a) and aliquots from a homogenized specimen of iron-loaded liver tissue. The coefficient of variation (CV) of LIC measurements between laboratories was 12 %, comparable with the CV of 11% found for interlaboratory LIC measurements in a previous study of 48 laboratories ⁹.

**R₂-LIC calibration**

An empirical analytical expression for a calibration curve relating liver <R₂> to biopsy LIC was found by modelling curves to the data with the aid of non-linear regression algorithms. The calibration equation has the form

\[ <R_2> = a + bx^d + cx^{2d}, \]

where <R₂> is the mean liver R₂ value in units of s⁻¹, x is the mean liver iron concentration in units of (mg Fe).(g dry tissue)⁻¹, and a, b, c, and d are constants with the
values, 6.88 s\(^{-1}\), 26.06 s\(^{-1}\),(mg Fe)\(^{-0.701}\).(g dry tissue)\(^{1.402}\), -0.438 s\(^{-1}\),(mg Fe)\(^{1.402}\).(g dry tissue)\(^{0.701}\), and 0.701 respectively.

**Statistical analyses**

The Spearman rank order test was used to determine the non-parametric correlation between the R\(_2\) measurements and liver biopsy LIC. The methods of Bland and Altman\(^{13}\) were used to determine the 95% limits of agreement between R\(_2\)-LIC measurements and biopsy-LIC measurements. Sensitivity and specificity of the R\(_2\)-LIC measurement to discrimination of biopsy LIC values above certain clinically important LIC thresholds were evaluated. Confidence limits for the sensitivity and specificity were obtained using the Wilson score method\(^ {14}\). Areas under receiver operating characteristic (ROC) plots were evaluated at each of the clinically important LIC thresholds by calculating the true positive fraction and true negative fraction for detection of LICs above the clinically important threshold for each possible cut-off value of mean liver R\(_2\)\(^ {15}\). Standard errors on the areas under the ROC plots were evaluated using the approximations of Hanley and McNeil\(^ {16}\).

**RESULTS**

**Measurement of R\(_2\) for aqueous MnCl\(_2\) Phantoms**

Measurements of R\(_2\) for a series of aqueous MnCl\(_2\) phantoms with a range of ionic concentrations were made using 13 different MR scanners (Fig 1) using techniques based
on methods described elsewhere \cite{10,12,17,18} and summarised earlier. The mean relaxivity value obtained from the thirteen scanners was 73.6 s$^{-1}$ (mM)$^{-1}$, with a standard deviation of 1.6 s$^{-1}$ (mM)$^{-1}$. The CV of relaxivity measured on the five scanners was 2.1 % demonstrating a high degree of reproducibility of the R$_2$ measurement technique on phantoms.

**Measurement of R$_2$ and iron concentration for liver tissue in vivo**

Biopsy LIC values measured for the 105 human subjects ranged from 0.3 to 42.7 mg Fe/g dry liver tissue. There was a highly significant correlation ($\rho = 0.98$, $P < 0.0001$) between biopsy LIC and mean liver R$_2$ ($<R_2>$) measurements for the region of interest in the right lobe of the liver as determined by the Spearman rank order test for all subjects (Fig 2). The 95% limits of agreement between R$_2$-LIC and biopsy LIC were found to be 50 and –56% (Fig 3). These limits of agreement are comparable with an expected repeatability coefficient between two needle biopsy LIC measurements from different parts of a fibrosis-free liver of 53% (based on an average CV of needle biopsy LIC measurements from a single liver of 19\% \cite{7,8} for biopsy specimens of less than 4 mg dry tissue).

The means and modes of the liver R$_2$ distributions shift to higher values of R$_2$ with increasing biopsy LIC (Fig 2 and Fig 4). Furthermore, there is a general increase in R$_2$ variability throughout the liver with increasing biopsy LIC, which is evident as a
broadening of the $R_2$ distribution (Fig 4 and Fig 5). The mean CV of $R_2$ throughout the maximum area axial slices of the liver for all 105 volunteers was $29 \pm SD 8\%$.

The sensitivity of liver $<R_2>$ measurement to biopsy LIC at low LIC values is illustrated in the inset of Figure 2. The sensitivities and specificities of the measured liver $<R_2>$ values for the discrimination of biopsy LIC values above various clinically significant thresholds $^{19,20}$ are shown in Table 1 along with the corresponding areas under the ROC plots. The calibration curve models the relationship between $<R_2>$ and LIC, with a Pearson’s correlation coefficient of 0.98 (Fig 2). Analysis of the data using the methods of Bland and Altman $^{13}$ showed that the mean differences between the $R_2$-LIC values and biopsy-LIC values for each of the individual subject groups were not significantly different from zero suggesting that the single calibration curve is sufficient to model the relationship between liver $<R_2>$ and LIC for all the subject groups. By removing all cases ($n = 14$) with a Knodell fibrosis staging of 5 or 6 $^{21,22}$, the mean difference between $R_2$-LIC and biopsy LIC was not significantly different from zero. A graphical representation showing the relationship between $R_2$-LIC, biopsy LIC and the scatter of data with different fibrosis stages is given in Figure 6. Of the 105 biopsies used in this study, only 48 had the dry masses permanently recorded by the pathology laboratory. Of these, 17 had dry masses below 1mg and 31 had dry masses greater than 1mg. By considering only biopsies with a recorded dry mass $>1mg$, the mean difference between $R_2$-LIC and biopsy LIC was not significantly different from zero.
No significant correlation was found between age and biopsy LIC or liver R$_2$ for the hemochromatosis subjects. However, among the β-thalassemia/Hb E subjects, there was a significant correlation between both age and biopsy LIC (Spearman rank order ρ = 0.37, $P = 0.02$) and between age and liver <R$_2$> (Spearman rank order ρ = 0.43, $P = 0.006$).

*Staging of fibrosis:* For the iron-loaded volunteer groups (HH, β-thal, and β-thal/Hb E), the CV of R$_2$ was weakly but significantly correlated with the staging of fibrosis according to the Knodell stage $^{21,22}$ ($\rho = 0.35$, $P = 0.003$, $n = 73$, Spearman rank test). For the HH patient group, there was a slightly stronger significant correlation between the CV of R$_2$ and the staging of fibrosis ($\rho = 0.48$, $P = 0.03$). Further, for the HH group, the skewness of the R$_2$ distribution was significantly correlated with the staging of fibrosis ($\rho = 0.52$, $P = 0.02$). Neither the CV of R$_2$ nor the skewness of the R$_2$ distribution were significantly correlated with LIC.

*Reproducibility study:* By measuring 10 subjects on two MRI scanners, we determined the random uncertainty on a single slice liver <R$_2$> measurement to be ± 7.7 %, with a non-significant systematic difference between the scanners of 1.2 % (with the systematic difference being less than 6.7% with 95% confidence).
DISCUSSION

Accuracy and precision of $R_2$ measurements

The accuracy and precision of $R_2$ measurements on thirteen 1.5 T MRI scanners have been demonstrated using aqueous MnCl$_2$ phantoms. The measured relaxivity value of $73.6 \pm 0.4 \text{ s}^{-1} \text{ mM}^{-1}$ is consistent with the value of $74.0 \text{ s}^{-1} \text{ mM}^{-1}$ from the measurements of Anderson and Jensen at 1.5 T (20 °C) $^{23}$. For the \textit{in vivo} measurement of liver $\langle R_2 \rangle$ values, the observed systematic difference between repeat liver $R_2$ measurements made on two MRI scanners was not greater than 6.7% (with 95% confidence), consistent with the measurement of a CV of MnCl$_2$ relaxivity between 13 different scanners of 2.1% for the phantoms. The random uncertainty on a single liver $\langle R_2 \rangle$ measurement \textit{in vivo} was approximately 8% while the coefficient of variation of $\langle R_2 \rangle$ for slices neighbouring the maximal cross-sectional slice for the 10 volunteers was 10%. Hence, it is possible that a significant part of the 8% random uncertainty is due to the naturally occurring variation in $R_2$ from slice to slice within each liver combined with inexact slice registration between the two measurements.

$R_2$ sensitivity, specificity, and dynamic range for LIC measurement

The specificity of liver $R_2$ (as measured by a variety of single spin echo techniques) for the quantification of liver iron concentration has been demonstrated in several previous studies of iron-loaded patients $^{17,18,24-26}$ using monoexponential signal decay analysis. We have demonstrated previously that variation of $R_2$ within a single liver reflects the LIC
variation throughout the liver\textsuperscript{17,18}. However, this is the first report to demonstrate a measurement method with (a) negligible instrument dependent systematic errors, (b) a universal calibration curve applicable to multiple patient groups with a variety of liver pathologies, and (c) iron concentration imaging capabilities. Furthermore, the calibration curve covers a greater dynamic range with a greater sensitivity and specificity of the measurement parameter ($R_2$ in this case) to needle biopsy measured LIC than any other reported MR methodology. High levels of sensitivity and specificity are demonstrated at the clinically important LIC thresholds of 1.8, 3.2, 7.0 and 15.0 mg Fe/ g dry tissue (see Table 1). At higher liver iron concentrations the sensitivity to needle biopsy LIC starts to drop presumably because of both the curvature in the relationship between $R_2$ and biopsy LIC and the increase in biopsy sampling error at higher LIC (e.g at 25 mg Fe/g dry tissue the sensitivity and specificity for biopsy LIC are 0.77 (95% confidence limits 0.50 to 0.92) and 0.98 (95% confidence limits 0.92 to 0.99) respectively with an area under the receiver operating curve of $0.67 \pm (SE 0.09)$.

The sensitivity and specificity of $<R_2>$-derived LIC measurement to biopsy LIC appears comparable with that reported for biomagnetic liver susceptometry (BLS), a technique which was first used clinically in the early eighties\textsuperscript{27}. BLS is reported to have an uncertainty of 50 – 300 $\mu$g Fe/ g wet tissue (or 0.17 – 1.00 mg Fe/ g dry tissue, assuming liver is 70% water\textsuperscript{28}) below approx 5 mg Fe/ g wet tissue (approx 16.5 mg Fe/ g dry tissue). Above 5 mg Fe/g wet tissue the differences between BLS LIC measurements and biopsy LIC measurements are greater and there is evidence for a departure from a linear relationship between BLS-measured LIC and biopsy-measured LIC which may be a
consequence of differences in the magnetic properties of ferritin and hemosiderin. Such phenomena may explain the curvilinear relationship between $<R_2>$ and biopsy LIC observed in the present study. However, other factors such as clustering of hemosiderin iron deposits at higher iron concentrations or a systematic change in tissue hydration with iron loading could also explain the curvilinear relationship between liver $<R_2>$ and LIC that we observe. Interestingly, a BLS study briefly reported by other workers shows noticeably smaller differences in BLS LIC measurements and biopsy LIC measurements at higher LIC values and the relationship between BLS LIC and biopsy LIC remains linear to high LIC values. The reasons for the better agreement between BLS LIC and biopsy LIC are not entirely clear but could include larger biopsy sample sizes, different pathologies, or more precise instrumentation.

Several reports have demonstrated that larger biopsy masses result in a more representative measure of average liver iron concentration or body iron stores. Very large biopsy specimens can result in lower coefficients of variation of LIC measurement. For example, Barry and Sherlock obtained a coefficient of variation of LIC in duplicate samples from remote sites of non-diseased liver of $8.6\%$ when using specimens with average dry mass of approximately $10$ mg. Angelucci et al. have shown that there is a better correlation between LIC measured by needle biopsy and total body iron stores when the sample mass is above $1$ mg. The pathology laboratories in our study routinely require a minimum biopsy dry mass of $0.4$ mg. Of the 48 biopsies that had their dry mass permanently recorded, 17 had dry masses below $1$ mg and 31 had dry masses greater than $1$ mg. The coefficients of variation of the differences in $R_2$-LIC and biopsy LIC were
25% and 22% for the biopsies with dry masses less than 1 mg and greater than 1 mg respectively, consistent with the observations of Angelucci et al. However, a Bland Altman analysis of the 95% limits of agreement between the LIC estimates by R2 measurement and biopsy were 39 ± 11% and −59 ± 11% for samples <1 mg, 46 ± 7% and −41 ± 7% for samples >1 mg, and 50 ± 5% and −56 ± 5% for all 105 biopsies in the study. With 17 samples below 1 mg and 31 samples above 1 mg, there was not enough power to detect a significant difference in the limits of agreement determined for the three groups of samples.

**Liver iron heterogeneity and R2**

Previous spin-echo imaging studies of the liver have shown correlation of R2 with liver biopsy iron concentration up to 22.4 mg Fe/g dry tissue using monoexponential signal decay analysis in selected regions of interest. Localised MR spectroscopy measurements in the vicinity of the biopsy site have shown good correlation (r = 0.95) with liver iron concentration up to 37.2 mg Fe/g dry tissue. However, a limitation of localised spectroscopy measurements is the inability to measure or image the spatial variation in liver iron concentration. Non-invasive liver iron measurement by biomagnetic liver susceptometry is similarly restricted to the measurement of LIC in smaller liver volumes.

Previous R2 imaging studies of iron loaded liver tissue post mortem have shown that the variation in R2 within a single liver reflects a spatial variation in liver iron concentration over a length scale of approximately 1 cm. The standard deviation of liver R2
distributions, $\sigma_{R2}$, appears to be a measure of the degree of heterogeneity in iron concentration throughout the liver. In this study, $\sigma_{R2}$ was found to be correlated with biopsy LIC. This observation is consistent with previous studies showing an approximately linear increase in the standard deviation of multiple site needle biopsy LIC measurements over a liver with the mean biopsy LIC measurement. For cirrhotic livers, the coefficient of variation of needle biopsy LIC has been reported to be 41% or greater, while in non-diseased liver average values of approx 19% have been obtained. Another study on multiple sampling of two non-cirrhotic post-mortem $\beta$-thalassemic liver tissue with larger samples (0.2-0.3 g) yielded coefficients of variation of LIC of 17 and 24%. Our finding that the CV of liver R$_2$ positively correlates with fibrosis stage in the iron loaded subjects further suggests that the CV of R$_2$ is a measure of the degree of heterogeneity of iron concentration within the liver, consistent with our previous finding that the spatial variation of R$_2$ within a single liver reflects the spatial variation of iron concentration within the liver. Hence further research is warranted to investigate whether spatial information generated from R$_2$ imaging may be used to assess the degree of liver fibrosis.

Liver iron concentration and age

No significant correlation was found between age and biopsy LIC or liver R$_2$ for the hereditary hemochromatosis subjects. This observation is consistent with a previous study of 410 subjects with hereditary hemochromatosis. The significant correlation between age and biopsy LIC (Spearman rank order $\rho = 0.37$, $P = 0.02$) and age and liver $<R_2>$ (Spearman rank order $\rho = 0.43$, $P = 0.006$) for the $\beta$-thalassemia/Hb E subjects is
most likely due to the fact that the patients had received no chelation therapy and few (if any) blood transfusions. In this group of subjects, iron loading is due to increased dietary absorption and so the correlation of LIC with age suggests a characteristic rate of iron uptake for the β-thalassemia/Hb E subjects.

Mechanisms of R₂ enhancement in iron loaded liver tissue

The mechanisms by which tissue iron deposits enhance proton transverse relaxation rates are not yet fully understood. However, two different mechanisms have been proposed, both of which may play a role in iron loaded tissue. Gossuin et al 35 have developed a theoretical model involving proton exchange between bulk water and exchangeable protons located at the surface of the hydrated iron(III) oxyhydroxide cores of ferritin. The model explains the magnitude of the effect of ferritin concentrations on the proton transverse relaxation rate in aqueous solutions. On the other hand, Jensen and Chandra 36 have developed a model that explains the non-exponential nature of proton transverse relaxation in iron-loaded tissue. Their model suggests a relaxation mechanism based on the diffusion of protons in the magnetic field inhomogeneities induced by micron scale hemosiderin clusters within the tissue. It is likely that both mechanisms play a role in liver tissue. As such, the relative fraction of iron in dispersed ferritin and clustered hemosiderin may influence the relaxivity of the tissue iron (i.e the amount of relaxation enhancement per iron atom). Thus a variability in this fraction may contribute to the variability in the relationship between hepatic R₂ and LIC and could be a possible explanation for the curvature seen in the R₂ vs LIC relationship in Figure 2. A variability
in the way iron clusters will also determine the relationship between $R_2$ and LIC as demonstrated by in vitro experiments with ferritin in liposomes. Spatial variations in these parameters throughout an individual liver could contribute to the variation in $R_2$ as shown in Figure 5.

**Other MRI methods of LIC measurement**

There have been several other MRI based methods for assessing liver iron concentration reported in the literature over the past two decades. They generally fall into four main categories: (i) signal intensity ratio methods based on $T_2$ contrast (e.g.), (ii) signal intensity ratio methods based on $T_2^*$ contrast (e.g.), (iii) relaxometry methods based on $T_2$ measurement (e.g.), and (iv) relaxometry methods based on $T_2^*$ measurement (e.g.). Signal intensity ratio methods generally enable shorter data acquisition times but are inherently less precise given that fewer data are acquired. The most promising of the signal intensity ratio methods is that recently reported by Gandon et al. For the purpose of distinguishing subjects with significant iron loading (defined by Gandon et al. as 60 $\mu$mol Fe/g dry tissue i.e approx 3.2 mg Fe/g dry tissue) from subjects without significant iron loading, their technique demonstrates a sensitivity of 89% and specificity of 80% (with somewhat higher values for their study group; sensitivity 93%, specificity 98%). Their technique enables measurements to be made up to LICs of 375 $\mu$mol Fe/g dry tissue (20.9 mg Fe/g dry tissue).
Of the T₂* relaxometry methods, the method of Anderson et al ⁴² has attracted the most attention recently. Anderson’s method was developed primarily for relaxometry measurements of the heart, and hence is not necessarily optimised for liver iron measurement. Nevertheless, a correlation between T₂* and liver iron concentration as measured by biopsy has been observed with a relatively good sensitivity and specificity for detecting liver iron concentrations above 3.2 mg Fe /g dry tissue (sensitivity approx 100% specificity approx 90%). However, sensitivity and specificity for discriminating LIC above other thresholds of clinical importance are less satisfactory (eg. a cut point of 7 mg Fe /g dry tissue gives approximate sensitivity and specificity of 70% and 88% respectively, while a cut point of 1.8 mg Fe/g dry tissue gives approximate sensitivity and specificity of 88% and 33% respectively). Subsequent developments of the technique of Anderson et al ⁴² have enabled T₂* relaxometry data to be acquired during a single breath-hold thus enabling very short data acquisition times ⁴³.

**Implementation of R₂-LIC Measurements in Clinical Practice**

The relative costs of the various methodologies will vary between institutions and countries. However, the overall costs will be determined by the length of time a patient spends inside the scanner and the time to analyse the data. The R₂ method described here involves a 20 minute data acquisition period (on average) compared with just a few minutes for single breath-hold methods. Nevertheless, 20 minutes is a relatively short period in the scanner compared with many other MRI examination protocols. The payoff for the extra minutes in the scanner is a higher specificity and sensitivity over a greater range of LIC than any other MRI based method of liver iron measurement. Data analysis
has been simplified by the production of custom designed software to facilitate the execution of the analysis algorithms.

The applicability of the technique has been demonstrated on a range of 1.5 T MRI units from the three major manufacturers. However, currently the short TE pulse sequences are not standard on the General Electric (GE) models.

The algorithms developed to measure and image $R_2$ in the liver $^{10,11,44}$ have been incorporated into a software package with associated training manuals. There are two general ways in which the software could become available to the medical community. Software could be distributed to individual MRI centres for use by individual radiologists. Alternatively, the telemedicine model could be used whereby data are transmitted to central data analysis facilities as a digital specimen to be analysed. The latter model has the advantages of quality control as well as eliminating the need for training of large numbers of radiologists.
ACKNOWLEDGEMENTS

We would like to thank the following people for supporting the research: Dr Jay Ives (SKG Radiology, Perth), Dr Catherine Cole (Princess Margaret Hospital, Perth), Dr John Pereira (Prince of Wales Hospital, Sydney), Enrico Rossi (PathCentre, Perth), Dr George Garas, (Sir Charles Gairdner Hospital, Perth), Pornpan Sirunkapracha (Mahidol University, Bangkok), Dr Pichet Metarrugcheep and Somsak Chanyawattiwongse (Neurological Institute, Bangkok).
REFERENCES


17. Clark PR, Chua-anusorn W, St Pierre TG. Proton transverse relaxation rate (R2) images of iron-loaded liver tissue: mapping local tissue iron concentrations with MRI. *Magnetic Resonance in Medicine*. 2003; 49: 572-575


<table>
<thead>
<tr>
<th>LIC Threshold</th>
<th>Clinical relevance</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Area under ROC plot</th>
</tr>
</thead>
<tbody>
<tr>
<td>mg Fe/g dry (µmol Fe/g DW)</td>
<td></td>
<td>(95% confidence intervals)</td>
<td>(95% confidence intervals)</td>
<td>(standard error)</td>
</tr>
<tr>
<td>1.8</td>
<td>Upper 95% of normal (^{19})</td>
<td>0.94</td>
<td>1.00</td>
<td>0.991</td>
</tr>
<tr>
<td>(32)</td>
<td></td>
<td>(0.86-0.97)</td>
<td>(0.88-1.00)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>3.2</td>
<td>Suggested lower limit of optimal range for liver iron concentrations for chelation therapy in transfusional Fe overload (^{20})</td>
<td>0.94</td>
<td>1.00</td>
<td>0.988</td>
</tr>
<tr>
<td>(57)</td>
<td></td>
<td>(0.85-0.98)</td>
<td>(0.91-1.00)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>7.0</td>
<td>Suggested upper limit of optimal range for liver iron concentrations for transfusional Fe overload and threshold for increased risk of iron induced complications (^{20})</td>
<td>0.89</td>
<td>0.96</td>
<td>0.991</td>
</tr>
<tr>
<td>(125)</td>
<td></td>
<td>(0.79-0.95)</td>
<td>(0.86-0.99)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>15.0</td>
<td>Threshold for greatly increased risk for cardiac disease and early death in patients with transfusional iron overload (^{20})</td>
<td>0.85</td>
<td>0.92</td>
<td>0.982</td>
</tr>
<tr>
<td>(269)</td>
<td></td>
<td>(0.70-0.94)</td>
<td>(0.83-0.96)</td>
<td>(0.016)</td>
</tr>
</tbody>
</table>

Table 1. The sensitivity and specificity of liver R\(_2\) measurements for discrimination of needle biopsy iron assay values above certain clinically important LIC thresholds are given together with their 95% confidence limits. The area under the ROC plot is given for each clinically important LIC threshold together with a standard error calculated by the method of Hanley and McNeil \(^{16}\) in order to give an approximate estimate of the uncertainty on the area.
FIGURE LEGENDS

Figure 1. \( <R_2> \) versus MnCl\(_2\) concentration for aqueous MnCl\(_2\) phantoms measured on thirteen different 1.5 T MR scanners. The solid lines are linear fits to the data.

Figure 2. \( R_2\)-LIC calibration curve. Liver \( <R_2> \) measurement for the region bounded by the surface of the right lobe of the liver and a sagittal plane 35 mm medial to the most lateral surface point versus needle biopsy liver iron concentration. The solid line is the calibration determined by curve fitting to the data. The Pearson’s correlation coefficient of the calibration with the data is \( r = 0.98 \). The error bars indicate the estimated ±19% uncertainties on biopsy measurement of average liver iron concentration. The uncertainty of 19% is based on studies of LIC heterogeneity in fibrosis-free liver \(^7,8\). The dashed lines indicate the 95% limits of agreement between \( R_2\)-LIC and biopsy LIC. Subject groups are ○ hepatitis, ■ hereditary hemochromatosis, ● β-thalassemia/Hb E, ♦ β-thalassemia.

Figure 3. Bland Altman plot showing the differences between \( R_2\)-LIC and biopsy LIC. The solid line shows the mean difference between the two measurements while the dashed lines indicate the upper and lower 95% limits of agreement between the two measurements. The different data symbols differentiate between the different fibrosis stages: stages 0 and 1, ○; stages 2 to 4, □; stages 5 and 6, ◊.
Figure 4. **Liver R₂ images and distributions.** Liver R₂ images and distributions for 4 subjects with different degrees of iron overload and pathological conditions (a) hepatitis (b) hereditary hemochromatosis (c) β-thalassemia (d) β-thalassemia/ Hb E. Note that the liver R₂ images are superimposed on standard spin-echo images for registration purposes. Principles of construction of R₂ images and distributions are described elsewhere. ⁴⁴ Note that to enable visualization of the heterogeneity of R₂ within each liver, the color-scale within each liver is adjusted for each image such that zero corresponds to voxel R₂ of zero while the maximum of the color-scale is scaled to the maximum R₂ value within the liver.

Figure 5. **Standard deviation of R₂ for the maximum area liver slice versus needle biopsy iron concentration.** The solid line is a linear fit to the data with a Pearson’s correlation coefficient of r = 0.89.

Figure 6. **R₂-LIC vs biopsy LIC.** The R₂-LIC values are derived from the calibration equation described in the text. The solid line is a straight line fitted through the origin and has a gradient of 0.980 ± 0.018. The different data symbols differentiate between the different fibrosis stages: stages 0 and 1, ○; stages 2 to 4, □; stages 5 and 6, ◊.
Figure 1.

Mean relaxivity = 73.6 s\(^{-1}\) (mM\(^{-1}\))

S.D. on relaxivity = 1.6 s\(^{-1}\) (mM\(^{-1}\))

CV = 2.1%
Figure 2.
Figure 3
Figure 4.
Figure 5.
Figure 6.
Non-Invasive Measurement and Imaging of Liver Iron Concentrations Using Proton Magnetic Resonance