# MEASUREMENT, MODELLING AND POTENTIAL CLINICAL APPLICATIONS OF SPATIAL VARIATIONS IN MAGNETIC RESONANCE PROTON TRANSVERSE RELAXATION RATES IN IRON-LOADED LIVER AND HEART TISSUE

---

This thesis is presented for the degree of Doctor of Philosophy of

The University of Western Australia

School of Physics

2006

Submitted by

Beau Pontré, B.Sc.(Hons)

I declare that this thesis is my own composition and that the work contained within has not previously been submitted for a degree at any educational institution

Beau Pontré

# Abstract

Magnetic resonance imaging (MRI) has been developed over the past two and a half decades to enable non-invasive assessment of soft tissues in the human body. MRI provides images of the tissues in the body with intensities weighted by nuclear magnetic relaxation properties of the tissue. Recent advances have utilised MRI as a quantitative tool with the nuclear magnetic relaxation rates in tissues being accurately quantified. One clinical application of quantitative MRI has been in the quantification of body iron stores in the management of iron overload diseases. MR images also contain information about the spatial variations of relaxation rates, which could be clinically useful. In the quantification of liver iron concentrations, proton transverse relaxation rate ($R_2$) maps have been used not only to quantify iron concentrations but also to visualise the spatial variations.

The work in this thesis addresses the use of spatial information from proton transverse relaxation rate maps in clinical practice. The quantitative spatial information contained in these maps is analysed in two clinically important settings, namely the non-invasive assessment of liver fibrosis and the assessment of magnetic susceptibility artefacts in cardiac proton transverse relaxometry.

Spatial distributions of liver $R_2$ maps were quantified using texture measures based on grey-tone spatial dependence (GTSD) matrices. Some of these measures gave a statistically significant distinction between patients with minimal or no fibrosis and those with fibrosis or cirrhosis. Distinction of fibrosis using this technique was enhanced in subjects with iron overload diseases, suggesting that iron is required as a contrast agent for sufficient sensitivity of image texture to fibrosis. In subjects with low tissue iron concentrations, tissue hydration was observed to also have an influence on $R_2$. In patients with end stage liver disease, a model combining tissue iron concentration and tissue hydration gave a better prediction of $R_2$ than iron concentration alone.

A model combining several of the texture measures was developed using logistic regression and was found to improve distinction of high-grade fibrosis from low-grade fibrosis. For the distinction of F0 and F1 fibrosis stages (as assessed by the METAVIR system) from F2 and above the area under the receiver-operating characteristic (ROC)

curve was 0.75. As this model was developed using a cohort of subjects with varying pathologies, the performance of the model is expected to improve if only iron-loaded subjects are considered.

Relaxation rate maps were also used to assess the magnitude and spatial extent of MR image artefacts near a magnetic susceptibility interface. The perturbation of the magnetic field by an object of a different magnetic susceptibility to a surrounding medium causes $R_2$* enhancement in the vicinity of the magnetic susceptibility interface. Results from experiments on simple geometric models of a sphere and cylinder were found to agree well with theory. The characteristics of the artefact at any point in a surrounding medium were found to be dependent on geometrical arrangement, size and distance away from the object causing the field perturbations.

In $R_2$* measurements of cardiac tissue, areas in the vicinity of major cardiac vessels are affected by magnetic susceptibility artefacts. The results obtained for average $R_2$* from maps of cardiac tissue provided similar results to clinically used region of interest (ROI) methods but provided the additional spatial information required to assess the artefacts. The magnitude of the artefact near the great cardiac vein (GCV) was found to correlate with septum $R_2$* ($\rho = 0.71$, $p = 0.0009$), suggesting that $R_2$* in the septum may be primarily determined by the magnetic susceptibility of blood in micro-vessels rather than iron deposited in the myocardium.

In summary, mapping of proton relaxation rates in tissue provides not only information regarding the mean proton relaxation properties of tissue but also spatial information. This spatial information can be both visualised and quantified to give additional information about the tissue, which may have some clinical utility. This spatial information has the potential to be applied to any number of clinical scenarios and is not limited to the applications investigated in this thesis.

# Table of Contents

# List of Figures

# List of Tables

# Acknowledgements

The work carried out in this thesis would not have been possible without the support of numerous people.

First and foremost, thanks must go to my supervisor Assoc. Prof. Tim St Pierre for his guidance throughout this project. Tim's understanding of biomagnetism and knowledge of experimental science was vital in all aspects of this research. His passion for this research and for science in general has been an inspiration. Thanks also to Dr. Wanida Chua-anusorn and Dr. Paul Clark who, along with Tim, provided the groundwork for this thesis with their research in the non-invasive measurement of liver iron concentrations with MRI. Their ongoing assistance has been invaluable for this research. Wanida's knowledge of biology and biochemistry was essential in the handling and analysis of biological specimens and in providing a clinical perspective to the research. Paul's expertise in programming was of great assistance in the development of the software for the texture analysis of the MR images and his help in the analysis of the liver MR images was essential for this research.

I am grateful for the support of the staff and students in the School of Physics at The University of Western Australia, particularly those in the Biomagnetics research group for providing a productive and enjoyable environment to work in. Thanks especially to Adam Fleming for his assistance with the liver transplant work and the MR data analysis and to those who helped out with the collection, transportation and analysis of the transplant samples. Thanks also to the School of Physics workshop staff for the construction and maintenance of essential laboratory equipment. Aside from the School of Physics, the Schools of Plant Science and Chemistry contributed to the preparation and analysis of tissue specimens.

Assistance and support for this research was also provided by a number of groups outside the university. This research was partly funded by the National Health and Medical Research Council of Australia. The liver transplant unit at Sir Charles Gairdner Hospital, Perth provided the tissue specimens for the study on excised liver tissue. In particular, thanks must go to Prof. Gary Jeffrey for assisting in the inception of the study and the recruitment of patients. Also, the help of Dr. Bastiaan DeBoer at

PathCentre, Sir Charles Gairdner Hospital was essential for the preparation of tissue specimens from liver transplant as well as the histological assessment of liver biopsy specimens.

The research in this thesis was centred on MRI and would have been impossible without access to the appropriate facilities. Dr. Jay Ives and the radiographers at SKG Radiology, St John of God Hospital, Perth provided imaging time and MRI facilities for much of the research in this thesis. Thanks also to Prof. Stephen Worthley at Royal Adelaide Hospital who provided the cardiac MRI facilities. I am especially grateful for the assistance of Greg Brown at Royal Adelaide Hospital for the collection of the cardiac MR images and for his part in the analysis of those images. Greg's advice on the physical principles of MRI and pulse sequences was invaluable not only in the cardiac MRI work but also in many other aspects of this research.

Throughout the duration of my PhD research, my family and friends have given me continuing support. Thanks especially to my parents, who have always encouraged me to pursue my goals and helped me to achieve them both during my PhD and throughout my lifetime. I am eternally grateful for their contribution but words cannot begin to express my gratitude. Finally, I would like to thank Clotilde for her untiring support, for without her, I would have had no reason to strive toward this endeavour.

# Chapter 1

# Magnetic Resonance Imaging as a Tool for the Non-invasive Assessment of Tissue Iron Concentrations and Soft Tissue Structure

Magnetic resonance imaging (MRI) is a clinical tool that can non-invasively image soft tissue in the body without the use of ionising radiation. Its operation is based on the principle of nuclear magnetic resonance (NMR), which was discovered independently by Bloch [1] and Purcell [2] in 1946 for which they were awarded the Nobel Prize in 1952. NMR was initially used as a spectroscopic technique for determining the physical and chemical composition of materials. Spectroscopy also allowed for the measurement of nuclear magnetic relaxation times, which are physical characteristics of given material. In 1971, Damadian [3] demonstrated the potential of NMR as a diagnostic tool, observing different relaxation times in tumour and normal tissue. Lauterbur demonstrated in 1973 [4] that images could be created using NMR by creating a two-dimensional image of test tubes using a back projection. The techniques for creating images using NMR were further developed by Ernst [5] who first used frequency-encoding, phase-encoding and the Fourier transform to create an image. In 1977, Mansfield presented an image of a cross-section through a finger [6], demonstrating the potential for MRI as a tool for non-invasive imaging. Lauterbur and Mansfield were awarded the Nobel Prize in medicine in 2003 for their contributions to the development of MRI. The first commercial scanners were developed in the early 1980's and have since been used widely in clinical practice.

In recent years there has been an emergence of clinical applications where relaxation times are quantified in MR images, rather than used only to provide a weighting to intensities in an image. In these applications, the accurate measurement of specific relaxation times are used as a non-invasive measure of some physiological effect and utilised as a diagnostic tool. One such application is in the quantification of body iron stores. Relaxation rates are strongly related to the iron concentration in tissue [7] and hence can be used as a non-invasive method of iron quantification. More recent developments in this application have enabled the accurate measurement of relaxation

rates in each voxel of an MR image of the liver [8]. The measurement of relaxation rates on a voxel-by-voxel basis gives information not only regarding the mean relaxation rates in the tissue but also information regarding the spatial distribution of relaxation rates.

The information contained in an MR image is not limited to the pixel intensities but also provides spatial information. For example, the spatial distribution of the intensities in the image provides information regarding the locations and sizes of various internal organs and regions of tissue, which are of diagnostic value. However, the assessment of spatial information extends well beyond the locating of tissue types and structures within the tissue. Quantifying the spatial information contained in these images could provide additional information regarding the characteristics of the tissue and enable a more accurate non-invasive assessment of the internal structure of organs or specific regions of tissue. The clinical value of these methods could be further enhanced when coupled with the accurate quantification of relaxation rates in tissue.

## *1.1 Iron Overload*

Iron is essential in biology as it is required in many metabolic functions. As a part of haemoglobin and myoglobin it facilitates the transport of oxygen throughout the body [9]. It is also required in other processes such as DNA synthesis [10] and electron transport [9]. However, iron in its free form is toxic as it acts as a catalyst to cell damaging reactions and hence needs to be stored and transported in a bound state [11]. Iron is stored within the protein ferritin to ensure that it remains non-reactive and the transferrin protein binds the iron so that it can be transported from storage sites to functional sites within the body.

Iron is absorbed through the diet in the mucosal cells of the gut. Homeostasis of iron is controlled solely by absorption, as there is no biological mechanism available to excrete iron. A healthy adult male has at any one time approximately 4g of iron [12]. On average, 1 – 2mg of iron is lost per day through sloughing of cells and other blood loss and in a healthy individual is accounted for by absorption of a similar quantity of iron through the diet [13]. The balance of iron metabolism and homeostasis can be disrupted by a variety of conditions. Iron deficiency can result from a number of factors where the body’s demand for iron exceeds the intake of iron from the diet [13]. Alternatively,

genetic disorders such as hereditary haemochromatosis and thalassaemia cause a build up of excess iron in the body.

Hereditary haemochromatosis is a genetic disorder that affects the regulation of iron absorption from the mucosal cells. The most common form is attributed to a mutation in the HFE gene [14]. In patients with hereditary haemochromatosis, the absorption of iron from the diet is about two to three times that seen in healthy individuals [13]. The excess iron is stored in the parenchymal cells of the liver and other organs [13]. As iron loading continues, liver fibrosis and cirrhosis develop and the patients are at an increased risk of cardiomyopathy owing to the iron deposition in the heart. The treatment of hereditary haemochromatosis is phlebotomy as blood letting in these patients causes increased erythropoietic activity requiring iron stored in the parenchymal cells to be mobilised.

Iron loading can also result from repeated blood transfusions, used in the treatment of some anaemias [15]. The most common of these, the thalassaemias, are a genetic disorder that results in the absence of deformation of the synthesis of α- and β-globin chains in haemoglobin. The β-thalassaemias are prevalent mainly in the Mediterranean, Africa and South-East Asia with the condition affecting up to 9% of the population in some regions [16]. Ineffective erythropoiesis in the thalassaemias causes iron accumulation through the absorption of excess iron from the diet [17]. However, when coupled with iron accumulation from repeated blood transfusions, severe iron loading develops and can cause irreversible damage in parenchymal tissues if left untreated. Treatment with iron chelation therapies is required to keep the body iron stores within acceptable limits [18].

In iron loading disorders, there is a need to measure the amount of iron in the body. Knowledge of the iron burden in a patient can be used to monitor iron chelation treatments and guide phlebotomy therapies [19]. As the liver is the largest store of iron in the body, a measure of liver iron concentration gives a good indication of body iron burden. The most direct way to measure liver iron concentrations is by chemical analysis of liver biopsy specimens.

## 1.2 Liver Biopsy

The utility of liver biopsy is not only in the measurement of liver iron concentrations but is also used to monitor the extent of liver fibrosis. In both of these applications, biopsy has proven to be clinically useful and remains the gold standard for measurement of fibrosis. However, there are a number of issues associated with biopsy that may make alternative methods more appropriate.

In both iron measurement and fibrosis staging, there is a sampling error in biopsy specimens owing to the small size of the specimen. A typical biopsy specimen is representative of about 1/50 000 the mass of the entire liver. Iron concentrations in the liver are not homogeneous and such a small specimen can introduce sampling errors. Studies have shown that the variations in iron concentration in normal liver tissue iron concentrations when sampled in specimens less than 4mg are typically 19% [20, 21] and can increase up to an average of 40% in end-stage liver disease [20, 21]. The processing of the biopsy specimen has been observed to introduce variability in the iron concentration measurement with the assessment of iron concentration from deparaffinized tissue samples overestimating that determined from fresh liver samples [22]. The measurement of fibrosis staging is also affected by sampling error of biopsy specimens. Evidence of cirrhosis may differ between the lobes of the liver [23]. The diagnosis of cirrhosis is incorrectly identified in 10 – 30% of patients with chronic liver disease [24]. One study found the number of false negatives from biopsy to be 24% [25]. Aside from the errors associated with the measurement of physiological parameters, biopsy is an invasive procedure and carries associated risks. In 1-5% of cases an extended stay in hospital is required owing to complications from the biopsy procedure [26, 27] and the procedure carries a 1/1 000 to 1/10 000 possibility of mortality [24].

When used for the assessment of liver fibrosis, the biopsy specimen is taken for histological analysis and assessed by a pathologist. The pathologist will assign a grading to the specimen indicating the level of fibrosis. This grading is usually based on one of two sets of criteria. The Knodell criteria [28] use a grading scale from 0 – 6 where 0 is allocated to a specimen with no evidence of fibrous tissue and 6 is assigned in the case of macronodular cirrhosis. Using the METAVIR scale [29] a grade of F0 – F4 is assigned with F0 being free of fibrosis and F4 assigned to specimens with cirrhosis. These methods have been used extensively in both a clinical and research

setting. However, these methods are highly subjective. Inter-observer and intra-observer variations can occur [30]. Ideally, a more objective measure of fibrosis is required. The assessment of any new measure needs to be compared with biopsy, which is not ideal owing to its subjective nature and susceptibility to sampling errors.

## 1.3 Liver Fibrosis

Liver fibrosis is a natural response to chronic liver disease. Conditions such as viral hepatitis, alcoholic liver disease and iron overload are among those that cause the accumulation of extracellular matrix (ECM) proteins. Liver injury resulting from chronic disease initiates reactions in inflammatory cells that release regulatory proteins to stimulate the production of the matrix. If the cause of the injury persists, the accumulation of the ECM proteins, such as collagen, will develop into fibrosis. The development of fibrosis is dependent on the underlying cause of the injury and different patterns of fibrosis emerge for different conditions [31]. If left untreated, the fibrous tissue can form cirrhotic nodules hindering the circulation in the organ rendering the liver incapable of performing its normal functions effectively.

In patients with chronic liver disease, fibrosis stage is an indicator of morbidity and mortality [32]. While liver biopsy remains the gold standard in fibrosis assessment, numerous non-invasive techniques have recently been developed and validated in clinical studies. However, serum markers, although sensitive, may not be specific for fibrosis as they can be affected by other factors [18]. There is evidence suggesting that the progression of fibrosis is reversible either by treatment or the removal of the cause of injury [33, 34]. A non-invasive method for measuring the degree of fibrosis in a patient will be useful not only for the on-going treatment of patients with chronic liver disease, but also for the evaluation and testing of anti-fibrotic treatments.

## 1.4 Non-invasive Alternatives

In recent years a number of alternatives to biopsy have been investigated as alternatives for both fibrosis and liver iron quantification. Techniques such as liver magnetic susceptometry and magnetic resonance imaging have been used for quantifying liver iron concentrations *in vivo* as well as serum tests for serum ferritin and trasnsferrin saturation. In fact, with all of these non-invasive alternatives, biopsy is not often used for these measurements. In the treatment of thalassaemia, the focus of body iron measurement has shifted to the heart recently. Heart failure induced by iron loading in

the myocardium is the most common cause of death in thalassaemia. The use of MRI in measuring heart iron concentrations has been used in the testing of oral iron chelators that have been claimed to remove iron from the myocardium.

## *1.5 Non-invasive Measurements of Fibrosis*

To date, the grading of liver biopsy specimens remains the gold standard for the assessment of liver fibrosis. A number of alternatives have been suggested and investigated. Serum tests and the analysis of biochemical markers have shown promising results as have a variety of radiographic techniques including ultrasound, elastography and MRI.

Liver biopsy is subject to sampling error resulting in cirrhosis being unidentified in up to 30% of cases [24]. Also, grading (using the Knodell or METAVIR criteria) is semi-quantitative and subjective in nature and hence is susceptible to intra- and inter-observer variability [30]. Additionally, the risks to patients do not allow frequent follow up studies and the risks coupled with the biopsy specimen size means that it is impractical to increase sample size through numerous biopsies. The development of a reliable method for the non-invasive assessment of fibrosis will allow for not only multiple follow up studies but also the chance to assess much larger regions of the liver tissue, greatly reducing the sampling error. Clinically, there is a need for a reliable non-invasive alternative [35].

Ultimately, all tests will have to be compared with the gold standard of histological grading of biopsy specimens. However, the subjective, semi-quantitative nature of this method may not be a reliable method for comparison. A number of studies have proposed more quantitative methods for the assessment of biopsy specimens. Pilette et al [36] found a good agreement between the semi-quantitative grading and the more quantitative area of fibrosis measure. However, the agreement with serum markers was higher when using the area of fibrosis measure than the semi-quantitative grading. Other methods utilising digital image processing have been proposed by [37-41].

### 1.5.1 Serum Tests / Biochemical Markers

Some markers in the serum have been associated with the deposition of ECM proteins. Measurements of these markers have to date provided some of the best alternatives to liver biopsy in the assessment of fibrosis. The most promising serum test, “FibroTest”

[42] uses five serum markers: alpha2-macroglobulin, haptoglobulin, gamma-glutamyl transpeptidase, bilirubin and apolipoprotein A1, as well as the inclusion of age and gender in the model. Initial results on hepatitis C patients reported a negative predictive value of 100%, positive predictive value of 91% and area under the receiver-operating characteristic curve (AUROC) of 0.87. Fibrotest has since been validated in alcoholic liver disease [43]. Using this test 46% of the patients would be able to avoid biopsy. FibroTest is a reproducible measure of fibrosis [44] and has a better five-year prognostic value than histology [45]. Although, Rossi et al [46] reported discrepant results in 11/57 (19%) of patients, Fibrotest performs better in discriminating fibrosis than historical indices [47], hyaluronic acid [43], transient elastography and glycomics [48]. Combinations of fibrotest with other non-invasive tests improve the discriminating power for fibrosis and cirrhosis [48]. A similar test “Hepascore” [49] uses sex, age and 4 serum markers (bilirubin, GGT, a2-macroglobulin and hyaluronic acid) in a model for predicting fibrosis. Hepascore performs best for the discrimination of advanced fibrosis (F3) with AUROC of 0.9 in a validation cohort and specificity and sensitivity of 74% and 88%. Other tests on biochemical markers [50-53] and DNA sequencing [54] have also proven to be indicative of fibrosis and cirrhosis.

Biochemical markers appear promising in the assessment of liver fibrosis but are not liver specific. All the tested markers can be affected by factors other than the development of liver fibrosis. Inflammatory activity resulting from the disease progression is related to many of the serum markers. Additionally, these markers will only increase significantly above normal in the initial onset of fibrosis and may not be useful in early detection [55]. Long term follow up studies are imperative to assess the efficacy of biochemical measures [56].

### 1.5.2 Transient Elastography

The level of fibrosis in the liver is related to the stiffness of the liver tissue. *In vitro*, liver elasticity measurements have been shown to correlate well with histological analysis for fibrosis [57]. Transient elastography “Fibroscan” allows an accurate, non-invasive and reproducible method for the quantification of liver elasticity [58]. This technique uses both ultrasound and low frequency waves to evaluate the physical properties of the liver tissue. The velocity of the low frequency wave is directly related to the elasticity of the tissue. Initial testing of this technique showed that the elasticity measurement correlated well with the fibrosis measure ($r = 0.71$, $p << 0.0001$). For the

distinction of fibrosis grades F2 and above, the AUROC was 0.88 and increased to 0.99 when used for the discrimination of cirrhosis (F4). The predictive power of FibroScan has been validated in other research [59] and found to perform better than individual serum markers [60].

### 1.5.3 MRI

A variety of different MRI sequences have been investigated in the assessment of liver fibrosis. In an animal model, a correlation between $T_2$ and connective tissue was observed but this was most likely a result of inflammation [61]. $T_1$ is significantly higher in cirrhotic patients than in controls according to a study by Thomsen et al [62], but no correlation with histological findings was observed. Superparamagnetic iron oxide (SPIO) and Gd enhanced $T_2$*-weighted imaging has also been investigated in the assessment of liver fibrosis [63]. The SPIO particles reduces the signal observed in regenerating nodules and the Gd makes fibrosis appear hyper-intense, greatly increasing the contrast between the nodules and fibrosis. Assessment of fibrosis from these images agreed well with fibrosis staging with a high sensitivity and specificity. Diffusion weighted imaging (DWI) has also shown promise for the discrimination of fibrosis [64] with a significant correlation observed between measures of cirrhosis and the apparent diffusion coefficient (ADC). However, DWI may not be sensitive enough to detect fibrosis in its early stages. Although a clinically robust method utilising MRI for the detection of fibrosis is yet to be developed, MRI methods provide the best option for fibrosis research [65].

### 1.5.4 Doppler Sonography

Doppler sonography has been investigated as a measure of fibrosis via blood flow velocity measurements in the liver. In the presence of fibrosis and cirrhosis, there is evidence that the fibrous tissue restricts blood flow in the major vessels in the liver. Portal vein velocity as measured by Doppler sonography is sensitive to hemodynamic changes owing to inflammation and fibrosis [66]. Using Doppler sonography, a diagnostic accuracy of 84% (75 – 90 95%CI) has been reported and has been used to correctly identify cirrhosis in 82 – 88% of patients [67]. Other parameters determined from Doppler sonography have also indicated sensitivity to fibrosis [68, 69]. However, these methods are subject to observer variability [67] and may not be appropriate for the staging of fibrosis [66].

## *1.6 Non-invasive Measurements of Body Iron Stores*

### 1.6.1 Serum Tests

Serum tests have been used to measure the degree of iron burden in a patient. Measurements of serum ferritin and serum transferrin saturation have been used for screening in hereditary haemochromatosis. Studies have shown correlations between liver iron concentrations and measures of serum ferritin. But these measures may not be reliable as serum ferritin can change for reasons not related to iron loading such as infection and inflammation [18].

### 1.6.2 Liver Magnetic Susceptometry

A number of radiographic techniques have also been investigated as a measure of iron concentration. While techniques based on computed tomography (CT) and nuclear resonance scattering (NRS) have been suggested and shown some degree of sensitivity and specificity for iron, they have not been adopted clinically owing to the use of ionising radiation [70]. Additionally, the density in the liver can be affected by factors other than iron making CT inappropriate in some pathologies. MRI has also been investigated in numerous studies and to date shows the most promise for being appropriate measure of iron concentrations.

The quantification of iron concentrations by magnetic measurement was initially suggested by Bauman and Harris [71]. More recently, devices based on the superconducting quantum-interference device (SQUID) magnetometer have been developed for the quantification of body iron stores [72]. The paramagnetic nature of iron deposits such as ferritin and hemosiderin makes SQUID magnetometry appropriate for body iron assessment as they dominate the magnetic response of tissue. While the response of paramagnetic species is on the order of $10^{-4}$ the strength of the applied field, the tissue itself, which is mainly diamagnetic, has a response of only about $10^{-6}$ of the applied field [72]. Additionally, ferromagnetic species are uncommon in human tissue and hence do not dominate the overall magnetic response.

Measurements of tissue magnetic susceptibility are made by moving the tissue of interest with reference to an applied external field. The changing magnetic flux induces a current in a detector coil. The SQUID sensor allows the detection of very small changes in flux and hence gives a very sensitive measure of the magnetisation of the

tissue. The complication to the measurement caused by the susceptibility interface between the tissue and air is accounted for by placing a water reservoir between the detector and the torso [73]. As the water has a similar magnetic susceptibility to that of iron free tissue, the measurements are not adversely affected.

The use of liver magnetic susceptomety for the quantification of liver iron has been investigated in a number of studies. Susceptometry methods are able to distinguish between normal and iron-loaded patients owing to the higher magnetic susceptibility in iron-loaded tissue [74]. In clinical use, very good correlations have been observed between magnetic measurements and liver iron concentration by Brittenham et al [72] ($r = 0.98$) and Fischer et al [75] ($r = 0.99$) up to 200μmol/g wet tissue. This method has also been used to assess the relationship between liver iron concentration and serum ferritin in thalassaemic children ($r = 0.64$) [76] and in other cohorts of haemochromatosis and thalassaemia patients [77, 78]. The efficacy of iron chelating therapies for the treatment of iron overload has also been observed using magnetometry methods [79, 80]. The utility of susceptometry may extend beyond the quantification of liver iron stores with recent studies investigating the use of susceptometry in other applications. One such application is the tracking of ferromagnetic tracers in the gastrointestinal tract [81] for the assessment of drug delivery [82]. Although the clinical utility of susceptometry has been proven, the limited availability of these devices worldwide may make this technology unsuitable for widespread clinical use.

## *1.7 MRI in the Quantification of Body Iron Stores*

A recent study in thalassaemia patients compared measurements of liver iron concentration as measured by liver biopsy, liver susceptometry and MRI and found that all three methods, although not equivalent, were correlated [83]. A correlation between liver susceptometry and MRI has also been observed in other studies [84]. The non-invasive nature of MRI, its high sensitivity to the paramagnetic iron deposits in tissue as well as its worldwide availability make MRI an ideal option for quantification of body iron stores in clinical practice. MRI methods have not only been used in the liver but also in other organs such as the heart and spleen where cellular iron deposition also occurs. Owing to its widespread use, MRI methods will be the focus of this thesis.

## 1.8 Magnetic Resonance Imaging

### 1.8.1 Theory

The nuclei of all atoms possess a property called spin and an associated angular momentum ($\vec{J}$). Although the spin is a quantum mechanical effect, classically it can be considered that a nucleus with non-zero spin will rotate about its axis [85]. As the nucleus is made up of charged particles, a loop of current is created giving rise to the nuclear magnetic moment, $\vec{\mu}$. The spin of a nucleus and the nuclear magnetic moment are related by the following equation

$$\vec{\mu} = \gamma \vec{J}$$

where $\gamma$ is the gyromagnetic ratio and is a constant for a given nucleus. Of the nuclei that are biologically abundant and have non-zero spin, hydrogen protons have the largest $\gamma = 2.675 \times 10^8$ rad $s^{-1}$ $T^{-1}$.

When placed in a magnetic field $B_0$, the magnetic moment of the nuclei will precess about the applied field direction at the Larmor frequency $\omega$, given by

$$\omega = \gamma B_0$$

The nuclear magnetic moment, as dictated by quantum mechanics, will exist in only a discrete number of orientations. For hydrogen, which has a nuclear spin of ½, there are only 2 discrete states [86]. A low-energy state where the z-component of the nuclear magnetic moment ($\mu_z$) is aligned parallel to the direction of $B_0$ and a high-energy state where $\mu_z$ is anti-parallel to $B_0$. These two states will be referred to as spin-up and spin-down respectively.

In an ensemble of protons, each of the individual nuclei will precess about the field direction in either the spin up or spin down orientation with random phase. At a finite temperature at thermal equilibrium, Boltzmann theory states that there will be a slight majority of spins in the spin-up state. The population difference in 1.5T at body temperature is about 1 in 200 000 [86]. The sum of the individual magnetic moments leads to a net magnetic moment $M_0$ that is parallel to $B_0$.

A radiofrequency (RF) pulse tuned to the Larmor frequency of the precessing nuclei will have a resonance effect. In addition to this, the individual spins will gain some phase coherence and precess about $B_0$ with the same phase. The energy imparted to the nuclei from the RF pulse is sufficient to cause transitions between the spin up and spin down state. The number of spins that undergo a transition will depend on both the duration and the intensity of the applied RF pulse.

When the RF pulse is applied to the extent where the populations of the spin-up and spin-down state become equal, this is known as a 90º or π/2-pulse. In this instance, the net magnetisation, owing to the phase coherence of the nuclei will precess about the applied field direction in the plane perpendicular to $B_0$. A 180º or π-pulse is when the RF pulse is such that the populations of the spin-up and spin-down state are inverted. If applied when the system is at equilibrium, the pulse will act to flip the net magnetisation anti-parallel to $B_0$.

Upon the removal of the RF stimulus from the system, the nuclei are allowed to relax back to the equilibrium state. This process happens via two distinct mechanisms [86]. Interactions with the thermal lattice of the system cause the nuclei to relax back to their equilibrium state, resulting in a recovery of the longitudinal magnetization. This process is known as spin-lattice relaxation and occurs with a characteristic spin-lattice relaxation time or $T_1$. Alternatively, each individual nucleus will experience a slightly different local environment resulting from interactions with neighbouring nuclei once the RF pulse is switched off. The nuclei then begin to precess at a Larmor frequency dictated by the local surroundings and slightly different from the other nuclei in the ensemble. The difference in local precession frequencies causes a loss of the phase coherence created by the RF pulse and consequently a decay in the transverse magnetisation. This mechanism is referred to as spin-spin relaxation and is characterised by the spin-spin relaxation time or $T_2$. The evolution of the net magnetisation with time is governed by the Bloch equations [1].

### 1.8.2 Measurement of Transverse Relaxation Rates

A measurement of the magnetisation can be made by applying a 90º pulse to the system and measuring the decay of the transverse magnetisation, sometimes known as the free induction decay (FID). In a homogeneous magnetic environment, this decay rate would

be equivalent to $T_2$. However, in an inhomogeneous magnetic field, like that experienced in an MRI scanner or in an inhomogeneous medium with magnetic susceptibility variations, macroscopic variations in the magnetic field environment cause a decay in the transverse magnetisation that is much greater than the than $T_2$ and is referred to as $T_2$*. To make a measurement of $T_2$, the decay in the transverse magnetisation due to the macroscopic field variations must be eliminated by the application of a spin-echo sequence [87].

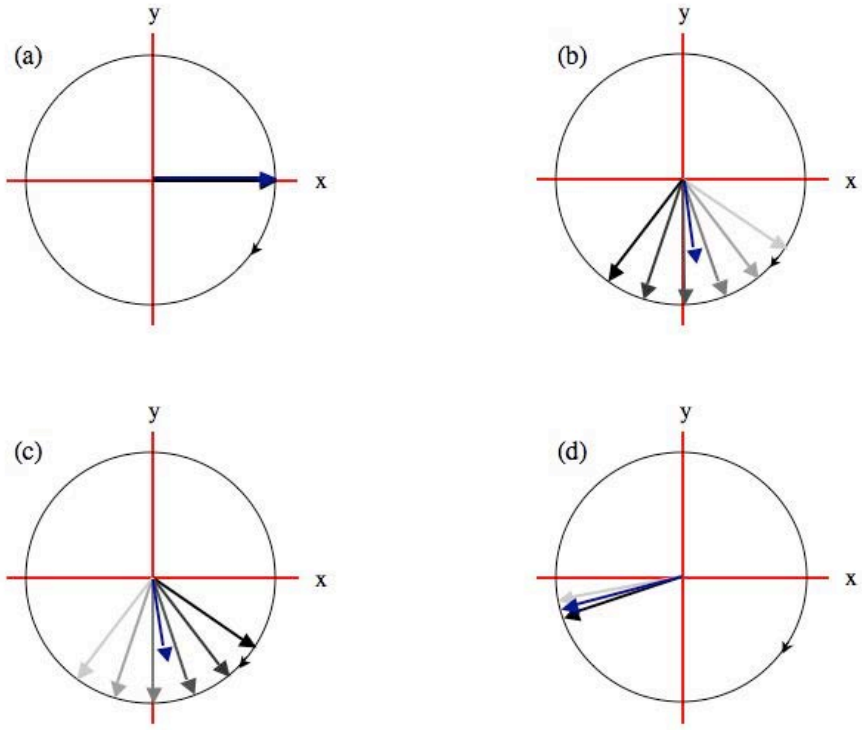



Figure 1-1 – The rephasing of spins with a spin-echo pulse sequence.

(a) The net magnetisation is rotated into the transverse plane with a 90º-pulse, (b) the spins begin to lose phase coherence as they experience different local field environments, (c) at time $\tau$, a 180º-pulse is applied to the system, reversing the phase of all individual spins, (d) at time $2\tau$ (echo time, TE) the spins rephase and form an echo. The blue arrow in each diagram represents the vector sum of the individual magnetic moments.

A spin-echo is achieved by the application of a 90º-pulse followed by a 180º-pulse. The 90º-pulse acts as described earlier and causes the net magnetisation to precess in the transverse plane (Figure 1-1a). Once the RF pulse is switched off, the spins start to dephase owing to both the spin-spin interactions and the interactions with the macroscopic field variations (Figure 1-1b). At a time, $\tau$, after the application of the 90º-

pulse, a 180º-pulse is applied to invert the phase of each nuclei, effectively putting the faster spins behind the slower precessing spins (Figure 1-1c). The spins then begin to rephase as the faster spins catch up to the slower ones. At time $2\tau$, the spins partially re-align causing an echo (Figure 1-1d). The re-phasing is only partial, as the dephasing that occurs as a result of the spin-spin interactions are not refocussed. The intensity of the echo signal is weighted by only the $T_2$ relaxation and the additional relaxation owing to field inhomogeneities does not contribute (Figure 1-2).

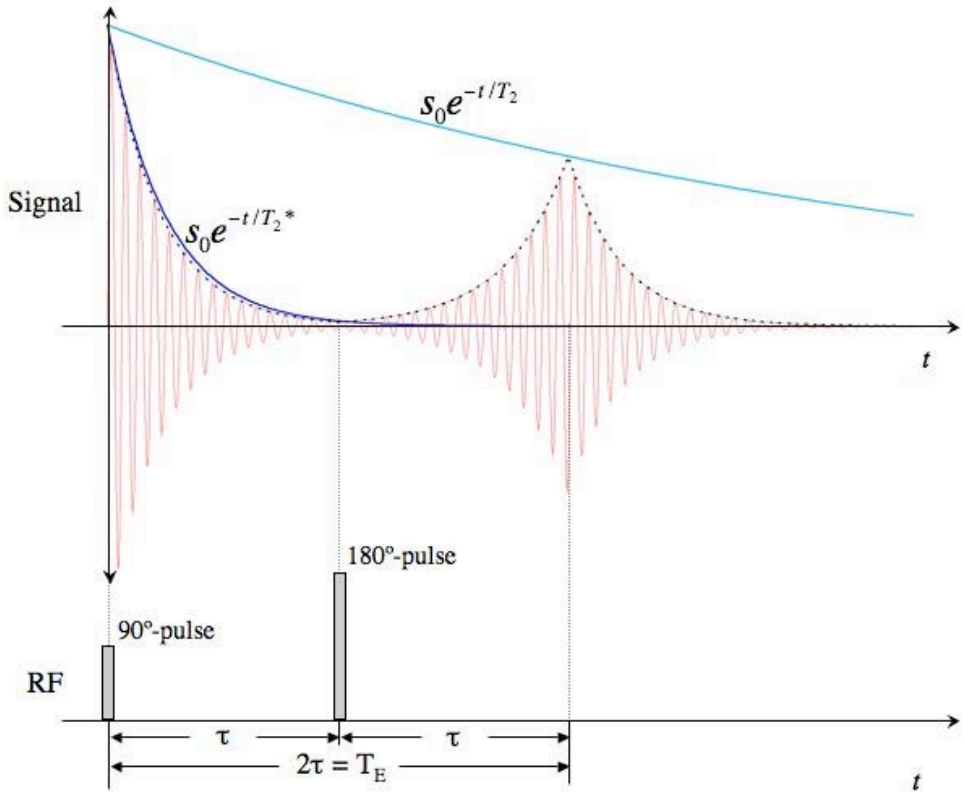



Figure 1-2 – Signal evolution during a spin-echo pulse sequence.

After the 90º-pulse, the signal decays with time constant $T_2$*, the application of a 180º-pulse refocusses the spins, forming an echo. The amplitude of the signal at TE is reduced from the intensity at time zero and is related to the $T_2$ decay envelope.

While a spin echo sequence gives a signal that is weighted purely by the effects of $T_2$, it has some practical limitations. With spin-echo sequences, the shortening of the repetition time TR leaves insufficient time for the recovery of the longitudinal magnetisation and leaves a smaller number of spins in the relaxed, equilibrium state. Hence, for the following 90º-pulse, there will be a decrease in the number of spins available for subsequent pulses and a decrease in the measurable signal. The application of a flip angle of less than 90º will leave a sufficient number of spins in the equilibrium

state for subsequent pulses but the 180º refocussing pulse will also invert the longitudinal magnetisation and saturate the spin system.

Gradient-echo sequences were developed to allow more rapid imaging while retaining sufficient signal for the measurements. In the interest of shorter scan times, an echo can be achieved by the reversal of the frequency-encoding gradient. The spins that were subject to slower precession frequencies precess faster after the gradient reversal and those spins that were precessing faster will now precess at a slower frequency. At the echo time, a signal can be detected. The reversal of the frequency-encoding gradient only allows the re-focussing of the effects attributable to the gradient itself. Unlike the spin echo case, the dephasing owing to the other field inhomogeneities are not refocussed and the intensity is weighted by $T_2$*, not $T_2$ [86].

The characteristic relaxation times $T_1$, $T_2$ and $T_2$* can be expressed as relaxation rates, $R_1 = 1/T_1$, $R_2 = 1/T_2$ and $R_2^* = 1/T_2^*$ respectively.

## *1.9 MRI in Measuring Liver Iron*

Relaxation times are affected by a number of effects. Different physical variations in tissue can affect $T_2$ and $T_1$ differently. Both are shortened in the presence of paramagnetic particles [88] with the effect on $T_2$ larger than on $T_1$ [89]. The higher sensitivity of $T_2$ to iron concentrations has been reported in both synthetic iron phantoms, ferritin solutions and iron loaded subjects [7]. As a result, the majority of studies into the use of MRI in iron quantification have involved either relative signal intensity measurements on $T_2$ and $T_2$* weighted images or the measurement of transverse relaxation times in the tissue of interest.

The methods used to calculate iron concentrations can be divided into two main groups. The first method involves the calculation of the signal intensity ratio between the tissue of interest and some reference tissue in the body (fat or skeletal muscle). The second type of analysis involves the quantification of the relaxation rates in the tissue. In early studies on liver iron quantification, many showed a much better correlation with iron using SIR methods over the relaxometry methods [90, 91]. The imaging parameters, in particular the echo times, used in these studies were not sufficient for the accurate quantification of $T_2$, especially in the presence of iron loading where the signal is within the noise floor of the image at long echo times. Studies that have used shorter echo

times have proven to be much more successful in the accurate measurements of $T_2$ [92, 93] and have shown better correlations with tissue iron concentrations. The accuracy of SIR methods has been improved by combining multiple regions of interest within the liver [94].

Gradient-echo (GE) methods for quantifying $T_2$* have been introduced [95] more recently and have a series of benefits. Firstly, the fact that a shorter TR is possible in GE sequences allows for images to be acquired within a single breath hold, which can greatly reduce motion artefacts. Secondly, the increased sensitivity to field inhomogeneities may increase sensitivity to hemosiderin deposits [96]. While the increased sensitivity causes much faster signal decay, GE methods are potentially useful up to similar iron concentrations as spin-echo based methods when sufficiently short echo times are used. However, the high sensitivity to field inhomogeneities can decrease the specificity of the relaxation measurement for iron since $T_2$* will be shortened by other macroscopic field inhomogeneities.

In the methods of St Pierre et al [8], the range of iron concentrations for which $R_2$ is appropriate has been extended by some technical considerations. Firstly, the use of spin density projection to estimate the theoretical TE = 0ms signal intensity gives more accurate quantification of decay curves where the signal reaches the noise floor at early echo times. Secondly, the use of a bi-exponential rather than mono-exponential decay model gave much better fits to the signal decay data at high iron concentrations. The introduction of these considerations has enabled the calibration of $R_2$ measurements with iron concentrations of up to 40mg/g dry tissue [97].

Most studies with MRI in iron overload have been performed on the liver owing to the fact that it is the organ where the majority of excess iron accumulates. Some investigators have also looked at iron deposition in the spleen, brain and heart. Johnston et al [98] suggests that iron deposition in the heart occurs after the accumulation of iron in the liver and spleen. A threshold effect of iron deposition in the heart has also been observed by Jensen et al, who reported in a group of non-thalassaemics with transfusional iron overload that myocardial iron becomes elevated only above a critical liver iron concentration of 350μmol/g [99]. Recently, a number of studies have investigated the use of myocardial proton relaxometry in the management of β-thalassaemia [95, 100, 101].

## *1.10 MRI in Measuring Heart Iron*

### 1.10.1 Importance of Heart Iron Measurement

Heart failure owing to excessive myocardial iron deposits is the most common cause of complication in thalassaemia accounting for 71% of patient deaths [102]. Iron deposits in the myocardium are not limited to thalassaemia and result from multiple blood transfusions in non-thalassaemics as well [103]. Early diagnosis is difficult as the clinical manifestations of iron-induced cardiomyopathy develop late. The symptoms include left ventricular disfunction and abnormalities in the echocardiogram [104, 105]. Cardiomyopathy is reversible if immediately treated with intravenous DFO [106, 107].

Measurement of cardiac iron concentrations may be the best indicator to assess which patients are at risk of heart failure. Early detection will allow treatment to be instituted early while the cardiomyopathy is still reversible [95]. Even in the absence of iron overload, evidence suggests that increased iron deposition in the myocardium could be related to cardiac disease [108].

An accurate method for the non-invasive quantification of cardiac iron is also essential for the testing of iron chelation therapies. The iron chelator deferoxamine (DFO) has been used in the treatment of iron overload diseases for over 30 years [109]. While DFO is clinically proven to reduce the mortality and morbidity in iron-loaded patients [110], compliance among users is low owing to the method of administration. Oral chelators such as deferiprone (L1) and Exjade (ICL670) are likely to improve the compliance of patients but there remains some controversy over the efficacy of L1. Using MR methods [111], L1 was reported to effectively remove iron from the myocardium but some studies have shown that liver iron remains high and in some instances above the threshold for cardiac disease [112, 113]. More recent research, however, has indicated that the use of L1 can improve ejection fraction and cardiac survival [114, 115]. It is apparent that the findings regarding the toxicity of L1 are conflicting and need to be investigated in further studies [109].

### 1.10.2 Methods of Measurement

Assessment of specimens obtained from endomyocardial biopsy is the only direct method for measuring myocardial iron concentrations *in vivo*. The specimens are usually analysed by histological staining but chemical analyses are also used.

Endomyocardial biopsy has not been used widely for iron quantification owing to its associated risks and sampling error. Variability of iron concentrations throughout the myocardium is considerable and hence biopsy specimens are subject to sampling errors [103]. A difference in iron deposition has been reported between the atria and ventricles [103] but also between the endomyocardium and epimyocardium where the iron concentrations are higher in the latter [116].

The only non-invasive method currently available for the quantification of heart iron is by MRI [96]. Although the number of studies in this area is limited, various different methods and pulse sequences have been investigated as a possible measure of heart iron concentrations. Animal models have shown that while $T_1$, $T_2$ and $T_2$* are all affected by the presence of iron in the heart, $T_2$ and $T_2$* measures are more sensitive and correlate better with myocardial iron concentrations [117-119]. Studies in humans have been mostly concerned with measures of $T_2$ and $T_2$* utilising both relaxomety and signal intensity methods with one study using images weighted by $T_1$ at low iron concentrations and $T_2$ at high iron concentrations [120]. This method was used to increase sensitivity of the method at low iron concentrations.

The first investigation into the measurement of cardiac iron concentrations was performed using measurements of signal intensity ratio at 0.5T [98]. While the SIR of liver to muscle and spleen to muscle were both found to be significantly higher in untreated haemochromatotics than in normals, no such relationship was observed in the SIR of the myocardium to muscle. At higher field strength [100] a significant difference in SIR was reported between thalassaemic patients and normal controls using a spin echo sequence (TE = 22ms). The same study also reported a correlation between SIR and serum ferritin ($r = -0.52$, $p < 0.001$). Jensen et al [120] used a method previously tested for the assessment of hepatic iron concentrations [121] based on a spin-echo method (TE = 25ms). Signal intensity measurements were made on the heart to muscle and liver to muscle in 41 transfusion dependent, non-thalassaemic patients and 15 controls. There was a significant difference in the heart SIR between the patients and the controls. The authors noted that the iron in the epimyocardium was significantly higher than the iron deposition in the endomyocardium in agreement with previous histological findings [116]. The heart iron measurement made by this method significantly correlated with serum ferritin ($\rho = 0.62$, $p < 0.0001$), liver iron

concentration, as measured by the same method ($\rho = 0.36$, $p = 0.02$) and the number of blood units given ($\rho = 0.45$, $p = 0.005$).

$T_2$ was measured in the myocardium for 40 thalassaemic patients and 9 controls (spin-echo TE = 22 and 60ms) [100]. A significant difference in $T_2$ was reported between the two groups. The same group in a subsequent study used pixel-by-pixel calculations of $T_2$ to assess a mapping method as a measure of myocardial iron [122]. These analyses were performed at 0.5T with TE = 12 – 120ms. A significant difference was observed in the mean $T_2$ values between the controls and thalassaemics with serum ferritin above 2000ug/L but not when the controls were compared to the thalassaemics with low serum ferritin. Additionally, there was a significant difference between the groups with high and low ferritin. The authors once again noted a correlation between heart $T_2$ and serum ferritin ($r = -0.67$, $p < 0.001$) and also reported a correlation between heart and liver $T_2$ for the thalassaemic patients ($r = 0.68$, $p < 0.001$). A correlation between serum ferritin and $T_2$ ($r = -0.932$, $p < 0.01$) was also observed as well as a significant difference in $T_2$ between the thalassaemia and control groups using a multiple spin-echo sequence (TE = 20 – 160ms) at a field strength of 1T [123]. Further studies in quantifying $T_2$ of the myocardium were carried out by Voskaridou et al [124], using a turbo spin echo sequence (TE = 5 – 80ms). A significant correlation between heart $T_2$ and serum ferritin in thalassaemia major patients was reported in this study. Heart $T_2$ values were also found to correlate with both liver $T_2$ and LVEF in the same group. Similar tests were carried out in sickle cell disease patients and thalassaemia intermedia patients but no correlations were observed. Owing to motion in the heart, $T_2$ measurements made by spin-echo may not be reliable as $T_1$ effects become significant for short repetition times.

A method using a gradient echo sequence to obtain $T_2$* values of the heart was developed for measuring cardiac iron [95]. In this study a series of images was acquired at multiple echo times between 5.6 and 17.6ms, each in a separate breath hold and using a mono-exponential fitting method to calculate $T_2$*. Below a $T_2$* of 20ms, a significant decrease in left ventricular function is reported and $T_2$* values above 20ms were found to coincide with the normal range of left ventricular function tests. No correlation was reported between liver and heart $T_2$*, in agreement with a later study [125]. However, many studies have reported a correlation between liver and heart proton relaxation parameters in iron-loaded patients [120, 122, 124]. In a subsequent study [126], a

multiple echo sequence produced similar results to the single gradient echo method. These measurements are also reproducible on different scanners [127].

The majority of studies investigating cardiac MR as a measure of heart iron do not measure cardiac iron concentrations directly. Most have either compared the measures of cardiac proton relaxation rates to either liver iron estimates [120], indirect measures of body iron burden such as serum ferritin [120, 123] or to tests of cardiac function [95]. Recently, Mavrogeni et al [101] performed a study investigating the relationship between cardiac $T_2$ and endomyocardial biopsy, carried out at low field strength (0.5T). A significant difference in cardiac $T_2$ between patients with low and high iron deposition was reported but there was no observed difference in the LVEF between the two groups.

Many of the methods described in this section have been used in the evaluation of iron chelation therapies. Anderson et al [111] observed the effects of both DFO and L1 and found that the L1 group had significantly lower cardiac iron (as assessed by cardiac $T_2$*) than the DFO group and had fewer patients within the abnormal $T_2$* range. They concluded that L1 is more effective than DFO at removing iron from the myocardium. However, Jensen et al [99] found that heart iron concentration, as measured by SIR methods, was normalised in 60% of patients with administration of DFO over 6-18 months. A study by Galia et al [128] using SIR measurements from a $T_2$-weighted GE sequence found that while both DFO and L1 gave changes in heart SIR (HSIR) within 1 year of administering treatment, only DFO acted to improve liver SIR. A later study by Anderson et al [129] saw an improvement in $T_2$* and cardiac function tests in 7 patients over 12 months of treatment with intra-venous DFO but reported that iron was cleared from the liver quicker than the heart. Caution must be taken in the interpretation of cardiac $T_2$* measurements as measures of iron as $T_2$* is yet to be calibrated with results from endomyocardial biopsy.

Although $T_2$* measures of the myocardium have proven to be clinically useful in the assessment and treatment of iron-overload, caution must be taken when interpreting these values as being indicative of myocardial iron concentration. Factors other than iron can cause changes in the $T_2$* measured in the myocardium. Artefacts owing to the location and geometry of major cardiac vessels have been reported to cause $T_2$* shortening in the vicinity of these vessels [130]. The microcapillaries in the

myocardium have also been reported to influence the $T_2$* values in the myocardium [131] and have been used at high field strengths as a means to visualise these capillaries [132]. A $T_2$* enhancement has been observed as a result of increasing blood oxygenation in these capillaries [133].

## 1.11 Assessing the Spatial Distribution of Relaxation Rates

To date, the majority of studies performed in MR-based iron quantification have utilised the calculation of relaxation rates based on the decay of the average intensity in a selected region of interest (ROI) [94, 95, 124]. While, these ROI methods have shown good correlations with mean liver iron concentrations and serum measures they only record a single value indicative of iron concentration in a small ROI in the liver. By averaging out the intensities, information regarding the spatial variations in the region of interest is lost. However, the distribution of iron throughout the liver is not homogeneous. A method for quantifying $R_2$ in liver tissue on a voxel-by-voxel basis has been developed to give maps of $R_2$ throughout the liver rather than just a single value [8]. The results from mapping techniques not only allow the quantification of mean $R_2$ but also an indication of the spatial variation of tissue iron within the liver.

The spatial variation of relaxation rates can be utilised to gain an understanding of the factors that cause the inhomogeneity. Factors such as varying tissue structure and certain pathologies could affect the distribution of relaxation rates. Additionally, the mapping techniques will allow the assessment of the spatial extent of artefacts in a given image quantitatively.

There are two situations of current clinical importance that can benefit from the application of mapping techniques. Firstly, the distribution of iron is affected by the presence of fibrous tissue in the liver [134]. Owing to the relationship between liver iron concentration and liver $R_2$, the spatial information in liver $R_2$ maps may be related to the extent of fibrosis or cirrhosis in the liver. Secondly, the application of mapping methods to the $T_2$* measurements in the heart for iron quantification may provide more information and allow the assessment of the artefacts that can complicate the images.

## 1.12 Thesis Aims

ROI methods of quantifying relaxation rates have been useful in the quantification of iron concentrations in a number of organs of clinical importance. Recent developments

in the field have allowed the voxel-by-voxel mapping of relaxation rates in human tissue *in vivo*. Mapping methods not only give information about the mean intensities in an image but potentially valuable information regarding the spatial distribution of relaxation rates. Spatial texture quantification has been used in many applications such as the assessment of Alzheimers disease [135] and the quantitative analysis of ultrasonic images of liver tissue [136], but has not been investigated in relaxation rate maps used in the quantification of iron concentrations. The mapping methods for accurately quantifying proton relaxation rates developed recently [8] allows for the investigation of the spatial distribution of relaxation rates in tissue. The aim of this thesis is to address the hypothesis that the assessment of spatial variations in proton transverse relaxation rates of human tissue from MRI is of clinical value. This investigation will be pursued with two applications in mind.

Firstly, as iron is preferentially accumulated in hepatocytes of the liver, the development of liver fibrosis is likely to introduce inhomogeneities into the iron distribution. In the case of macronodular cirrhosis, the appearance of the iron also becomes nodular [134]. It is proposed that these inhomogeneities not only affect the distribution of iron on the microscopic level but also influence the distribution on the macroscopic scale. The use of texture quantification on liver $R_2$ maps will be investigated to as a possible method for the non-invasive assessment of liver fibrosis.

Secondly, it has been previously observed that in cardiac imaging, cardiac vessels can cause artefacts in the myocardium. These artefacts are of most relevance in $T_2$* imaging where the relaxation rate is sensitive to inhomogeneities in the macroscopic magnetic field. $T_2$* methods are currently of clinical importance in the quantification of myocardial iron and the treatment of iron overload disease. The use of relaxation rate mapping methods will also be used to quantify both the intensity and spatial extent of these artefacts and the potential implications on the quantification of myocardial iron concentrations.

## 1.13 Thesis Outline

This thesis will be presented in two main sections. The first (Chapters 2 – 5) will investigate the possibility for spatial characterisation of $R_2$ maps to be used as an indicator for liver fibrosis. The second (Chapters 6 and 7) will address various methods

of measuring $R_2$* in the myocardium and the effects of magnetic susceptibility artefacts on the accuracy of these calculations.

Chapter 2 will introduce the methods that will be used throughout this thesis for the analysis of MR images. A description of the relaxometry mapping methods will be given and are central to all the analyses in the chapters to follow. A method for the texture analyses that will be used to quantify the spatial distributions of the relaxation rate maps will also be described. The texture analyses will also be applied to a series of test images to assess the appropriateness of such measures for the analysis of relaxation rate maps.

Chapter 3 presents the effects of texture analysis applied to liver $R_2$ maps *in vivo*. The spatial variations of $R_2$ in the liver tissue will be quantified using the methods of Chapter 2 and these results compared to fibrosis staging in the patients. The relationship between the spatial distribution of $R_2$ and fibrosis staging will be addressed in both the entire cohort and in sub-groups of patients with similar disease characteristics. An assessment of the reliability of these methods and an investigation into factors related to image analysis and acquisition that may complicate these measurements will also be presented.

Chapter 4 presents an attempt to use the relationships between texture and fibrosis observed in Chapter 3 to propose a model for the non-invasive prediction of liver fibrosis. A candidate model will be developed using statistical means. An assessment of the clinical applicability of this model as well as an investigation into the reproducibility and robustness of this test will be presented.

Chapter 5 presents a study on excised liver tissue from non iron-loaded patients with end stage liver disease. The specimens will be used to assess the efficacy of $R_2$ as a measure of iron concentration in non-iron loaded patients with end stage liver disease. This relationship will be compared to an existing calibration between liver iron concentration and $R_2$. The effect of tissue hydration on $R_2$ in these specimens will also be investigated. Variations in $R_2$ and liver iron concentration throughout the liver will also be evaluated. These assessments will provide insight into the use of texture characterisation on $R_2$ maps as an indicator of fibrosis in the presence of low iron concentrations.

Chapter 6 presents an investigation on synthetic phantoms to quantify the magnetic susceptibility artefacts that can confound measurements of $R_2$* in relaxometry techniques. Various geometries will be investigated to provide insight into the nature of these artefacts and how they compare with theoretical expectations. The consequences of magnetic susceptibility artefacts in cardiac MR imaging for the quantification of myocardial iron will also be addressed.

Chapter 7 presents a study on thalassaemia patients to compare various methods of measuring $R_2$* in the cardiac septum. The effects of echo time selection and decay models will be assessed as well as a comparison of mapping and ROI based methods of analysis. These results will be used to provide information about the variability observed in relaxation rate calculations with different analysis methods. The use of mapping methods, while similar to ROI methods provides additional information and can be used to investigate magnetic susceptibility artefacts that can complicate *in vivo* $R_2$* measurements in cardiac tissue.

Chapter 8 presents the general conclusions of this thesis and potential directions for future investigation.

# Chapter 2

# Methods of Quantifying Proton Transverse Relaxation Rates and Their Spatial Distribution

The work in this thesis is based on the assumption that MRI can provide information regarding the spatial variations of proton relaxation rates. To acquire this information, a method for accurately quantifying and mapping relaxation rates is required. In the quantification of body iron stores, most researchers have adopted analysis methods that measure the average signal decay in a given region of interest. While some of these methods have been of clinical value, they are not appropriate for this thesis investigation, as they do not provide information about the spatial variations of relaxation rates. A method for mapping the proton transverse relaxation rate ($R_2$) in liver tissue has been developed recently [8] and tested in comparisons with liver biopsy [97]. The methods will be used throughout this thesis to generate relaxation rate maps in tissue and in phantoms.

With a method to accurately calculate relaxation rates on a voxel-by-voxel basis, spatial variations can be visualised. However, the relaxation rate information can be used to obtain a quantitative description of the distribution and texture of relaxation rates in the image. Although there is a vast number of methods available to quantify spatial variation in digital images, the methods described by Haralick et al [137] will be used to quantify the texture in the relaxation rate maps throughout this thesis.

This chapter will describe the details of the acquisition and analysis of MRI data, based on the methods developed by St Pierre et al [8], as well as a description of the texture quantification methods as defined by Haralick et al [137].

## *2.1 Magnetic Resonance Imaging Methods*

### 2.1.1 Image Acquisition

In the method originally described by Clark et al [138] and later developed for the quantification of liver iron concentrations [8], MR scans were performed on a subject

using a series of single spin-echo sequences. In initial studies, axial images of the liver were acquired with echo times of 6, 9, 12, 15 and 18ms with a repetition time of 2500ms and 5mm slice thickness. The field of view was allowed to vary and was chosen according to the size of the patient. To increase the sensitivity of this method to high iron concentrations, echo times of 7 and 8ms were added to the protocol to give a better description of the early part of the decay profile. A bag of Hartmann's solution is positioned next to the patient in each scan. The solution acts as a reference standard to correct for any changes in the gain settings that may occur between each image acquisition, owing to its relatively long $T_2$.

The MRI experiments carried out in this thesis require different types of sequences to be implemented for a variety of specimens. As a result, no standard method of image acquisition will be used for all studies. All will involve the collection of a series of images at different echo times to allow for the calculation of relaxation rates in each specimen, the details of which will be described separately for each study. Regardless, the image analysis methods to be described in the next section can be applied regardless of the methods used to obtain the images.

### 2.1.2 Image Processing and Analysis for Generating $R_2$ Maps

All *in vivo* scans for the quantification of liver $R_2$ using the spin-echo sequences are analysed by the method of St Pierre et al [8]. This analysis method takes into account and corrects for a number of factors that may perturb the measured $R_2$ from the real values, such as gain drift, very high iron concentrations, background noise, and breathing artefacts.

To correct for gain drift between the image acquisitions a bag of Hartmann's solution is used. As the solution has a long $T_2$ as compared to the longest echo time used in the data acquisition, the solution acts as a reference standard in each image. The intensity of each voxel in the images is then scaled according the mean intensity of the Hartmann's solution in the 6ms spin-echo image.

Background noise arises in MR images owing to thermal effects and can be modelled by a generalised Poisson distribution. To correct for the noise, a region of interest is selected in a region of free space in the image unaffected by motion artefacts. A fit of the Poisson model to this region of interest gives rise to a signal offset that is

characteristic of the noise in the image. The signal offset is subtracted from the measured signal of each voxel in the image to give a corrected signal intensity for each voxel. Owing to the complex nature of the noise distribution, the subtraction of the signal offset from the measured signal must be performed in quadrature.

Long repetition times are utilised in this method to allow sufficient time for the recovery of the longitudinal magnetisation between successive spin-echo pulses. The major draw back of this is that image acquisition times are long and the images are susceptible to artefacts owing to patient motion. One part of this methodology is the correction of these artefacts post-acquisition. While the details of this process will not be described here these corrections have been shown to reduce the motion artefacts and enhance the accuracy of results obtained from $R_2$ mapping [139]. Other variations in intensity owing to image noise are smoothed by averaging neighbouring voxels within a 5 x 5 window kernel.

To achieve an improved fit to the decay profile, a method is used to estimate the initial proton spin density in the region of interest. The initial spin density coupled with a description of the RF field intensity in the image allows the calculation of the theoretical signal intensity for TE = 0ms in each voxel of the region of interest. The inclusion of the extra data point at TE = 0ms gives a better description of the decay profile for early echo times. An accurate description of the early part of the decay is especially important in the case of severely iron-loaded liver where there is minimal signal in the liver at longer echo times.

Once the pre-processing of the images is performed and the necessary corrections are made to the images, the relaxation rates can be measured. The relaxation rates are calculated on a voxel-by-voxel basis by fitting a model to the decay profile of the corrected signal intensity with increasing echo time. In the analysis of the phantoms in this thesis, a decay based on a single compartment model is expected. So for the analysis of phantoms it is expected that the signal will decay according to the equation [140]

$$S(TE,TR) = N(H)(1 - 2e^{-\frac{TR-TE/2}{T_1}} + e^{-\frac{TR}{T_1}})e^{-\frac{TE}{T_2}}$$

However, in the case of long repetition time as compared to the $T_1$, the equation reduces to

$$S(TE) = S(0)e^{-\frac{TE}{T_2}}$$
$$= S(0)e^{-R_2.TE}$$

where $S(0)$ is the signal intensity at TE = 0 and TE is the echo time.

In tissue however, a compartmentalisation of protons results from intracellular and extracellular water [141]. Owing to this compartmentalisation a multi-exponential decay model is applicable, rather than the single exponential decay observed in homogeneous media. A bi-exponential model has been sufficient to give reliable fits to relaxometry data in liver tissue [142] and is given by

$$S(TE) = S_f(0)e^{-R_{2f}.TE} + S_s(0)e^{-R_{2s}.TE}$$

The numerical methods utilised to fit the data to this model and calculate the fast and slow $R_2$ components are described elsewhere [8]. The average relaxation rate is determined from these compartments by

$$R_2 = \rho_f R_{2f} + \rho_s R_{2s}$$

where $\rho_f = \frac{S_f(0)}{S_f(0) + S_s(0)}$ and $\rho_s = \frac{S_s(0)}{S_f(0) + S_s(0)}$.

Throughout this thesis, unless stated otherwise, the relaxation rate maps will be determined using these relaxometry methods. An example of a liver $R_2$ map is given in Figure 2-1.

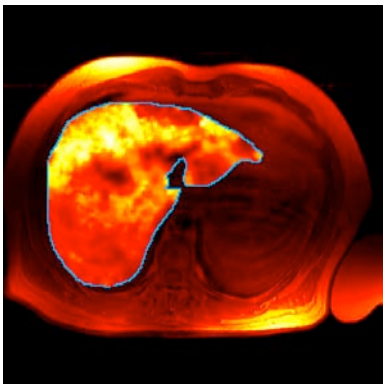


Figure 2-1 – Example of a liver $R_2$ map superimposed on a false colour spin-echo image.

Although the methods described in this section are used for the quantification of $R_2$ in liver tissue, these methods are not specific to a single application. Other relaxation rates such as $R_2$* can be determined by applying these image processing methods to gradient-echo images rather than spin-echo images. Additionally, the same methods can be used to measure relaxation rates in various tissues both *in vivo* and *in vitro* as well as in synthetic phantoms. Furthermore, the generation of both $R_2$ and $R_2$* maps in the same region of interest allows the generation of an $R_2'$ map by pixel-wise subtraction of the two images according to the equation $R_2' = R_2 * - R_2$.

## 2.2 Quantifying Texture in Liver $R_2$ Maps

At present, the clinical use of proton relaxometry techniques lies in the ability to quantify non-invasively mean $R_2$ and hence mean iron concentration. However, there is much more information to be obtained from these techniques. A map of $R_2$ contains spatial information as well as intensity information. It is apparent from this map that there is an inhomogeneous distribution of $R_2$ and hence iron. Some of these variations appear to be on the order of a few pixels while other features in the image may be up to 20 pixels or more in size (Figure 2-1).

This apparent inhomogeneity in the $R_2$ and iron distribution could be due to numerous factors, some of which will be addressed in a later chapter. Whatever the reason for these inhomogeneities in the $R_2$ maps, some method for assessing and quantifying this distribution is required to make comparisons with other observations.

There are many techniques available to assess and quantify texture and pixel intensity distributions in digital images. Simple techniques such as measurements of the standard deviation of pixel intensities in an image are easy to perform and give a numerical indication of the distribution of intensities in an image. However, there are many other, more complex methods that give a more descriptive measure of the texture in an image. Numerous methods of texture quantification have been used in the analysis of radiographic images for various applications [143].

While any of these measures could be used for the texture quantification of liver $R_2$ maps, the work in this thesis will concentrate on a particular class of texture quantification methods. Texture measures derived from grey tone spatial dependence (GTSD) matrices [137] have been used in a variety of applications including satellite imagery classification [137] and medical applications such as the assessment of Alzheimer's disease from MR images [135]. The mathematical relationship between texture measures calculated from GTSD matrices and histograms and the ability to probe texture variations on various length scales makes it appropriate for use on liver $R_2$ maps.

### 2.2.1 Grey-tone Spatial Dependence Matrices

A grey-tone spatial dependence (GTSD) matrix is representative of the texture in an image as each matrix element (*i*,*j*) represents the probability that a pixel of intensity *i* and a pixel of intensity *j* are neighbours. To illustrate the construction of a GTSD matrix, take the example of a 4 x 4 pixel image as in [137], with 4 grey levels - 0, 1, 2 and 3 (Figure 2-2).

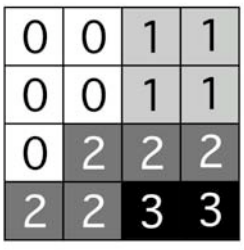



Figure 2-2 – Example of a 4 x 4 pixel image with 4 distinct grey levels (0-3).

Each element (*i*,*j*) of the GSTD matrix is the number of times that a pixel of intensity *i* is nearest neighbours with a pixel of intensity *j*. As there are 4 grey levels, the GTSD matrix will be a 4 by 4 matrix. Choosing different criteria for defining a pair of

neighbouring pixels will give different GTSD matrices. For example, if horizontally neighbouring pixels are considered (Figure 2-3a) the GTSD matrix will be different from that found if only vertically neighbouring pixels are used (Figure 2-3b). As pixels are considered neighbouring on both sides of the specified direction, these matrices are always symmetric about the diagonal.

$$P_H = \begin{bmatrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} \quad \text{(a)} \qquad P_V = \begin{bmatrix} 6 & 0 & 2 & 0 \\ 0 & 4 & 2 & 0 \\ 2 & 2 & 2 & 2 \\ 0 & 0 & 2 & 0 \end{bmatrix} \quad \text{(b)}$$

Figure 2-3 – GTSD matrices derived from the sample 4 x 4 pixel image in Figure 2-2

Determined for (a) horizontally neighbouring pixels, (b) vertically neighbouring pixels.

This technique can easily be extended to any direction and can also be used for various pixel separations. So instead of looking at adjacent pixels, the next nearest neighbours could be considered. Theoretically, a GTSD matrix could be constructed for any pixel separation within the physical size of the image of interest.

Although these parameters can be determined by comparing pixels in various directions, there is no reason to expect the $R_2$ to exhibit any directional texture features. Owing to the lack of directional dependence, the texture measures used will be defined at each pixel separation as the average of the given parameter over four directions – horizontal, vertical, left diagonal and right diagonal (Figure 2-4).

For the special case where the pixel separation is 0, all off-diagonal elements in the GTSD matrix will be 0 as the “neighbouring pixel” will always be the same intensity as the pixel of interest. In this case, the elements along the diagonal will represent a histogram of the $R_2$ values in an image. So this texture characterisation technique can be considered as an extension of texture measures calculated from a histogram. It has been suggested that the characteristics of the histogram of $R_2$ distribution, such as the coefficient of variation, in a liver are indicative of physical properties of the tissue such as liver fibrosis [97]. Hence texture measures derived from GTSD matrices could be appropriate for this application.

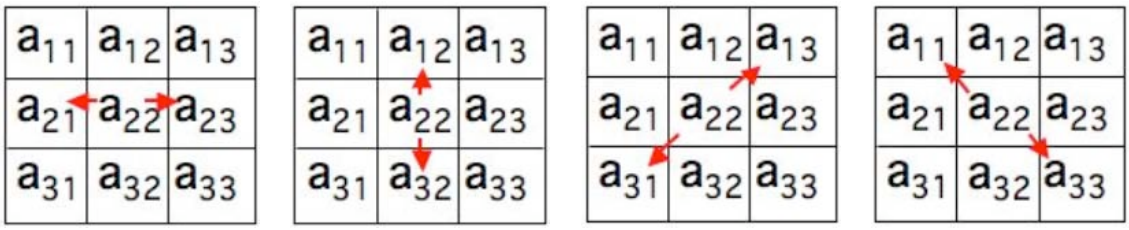



Figure 2-4 – Graphical representation of the four directions of neighbouring pixels

(a) horizontal, (b) vertical, (c) right diagonal, (d) left diagonal.

### 2.2.2 GTSD Matrices on Liver $R_2$ Maps

Take the liver $R_2$ map in Figure 2-1 an example. A histogram can be constructed from the intensity values in this map as shown in Figure 2-5.

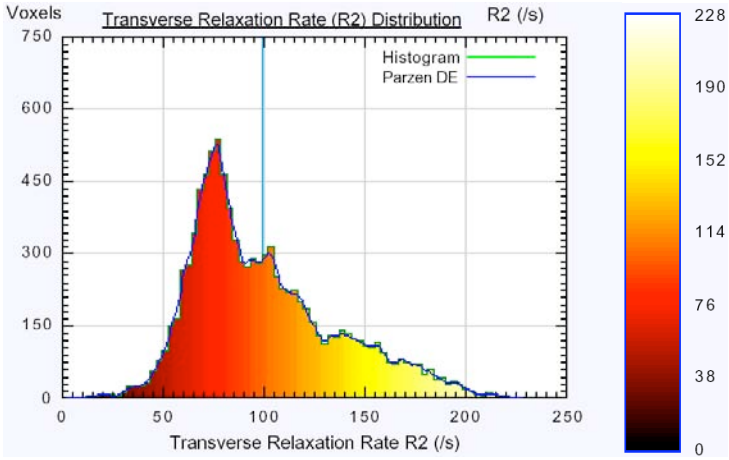



Figure 2-5 – Histogram of the distribution of $R_2$ values throughout the slice of liver in Figure 2-1.

Prior to the generation of GTSD matrices, the intensities in the liver $R_2$ maps are quantised. For each liver $R_2$ map the optimal bin width for quantisation ($q_w$) was determined by the expression $q_w = 3.49\sigma N^{-1/3}$ [144], where $\sigma$ is the standard deviation of the distribution of intensities and $N$ is the number of pixels in the liver $R_2$ map. As the size and construction of the matrix is dependent on relative pixel intensities on the quantised image, there are no restrictions on the size or shape of the region of interest. Freedom in the choice of geometry of the ROI allows the whole of $R_2$ map to be included in the analysis and hence the analysis is more representative of the patterns over the whole slice rather that just a smaller region of a specific geometry. As the pixel separation increases, the number of pixels that can be included in the construction of the

GTSD matrix is reduced, creating a band of pixels around the edge of the liver that cannot be included in the GTSD construction.

Additionally, the GTSD matrices can be represented graphically (Figure 2-6). The intensities on the diagonal of the GTSD matrix for d = 0 are representative of the counts for each bin in the histogram. By looking at non-zero pixel separations, information not contained in the histogram alone becomes available for analysis. For small pixel separations (eg d = 1) only those elements that are close to the diagonal remain non-zero. The spread of the non-zero matrix elements away from the diagonal becomes larger as the pixel separation increases.

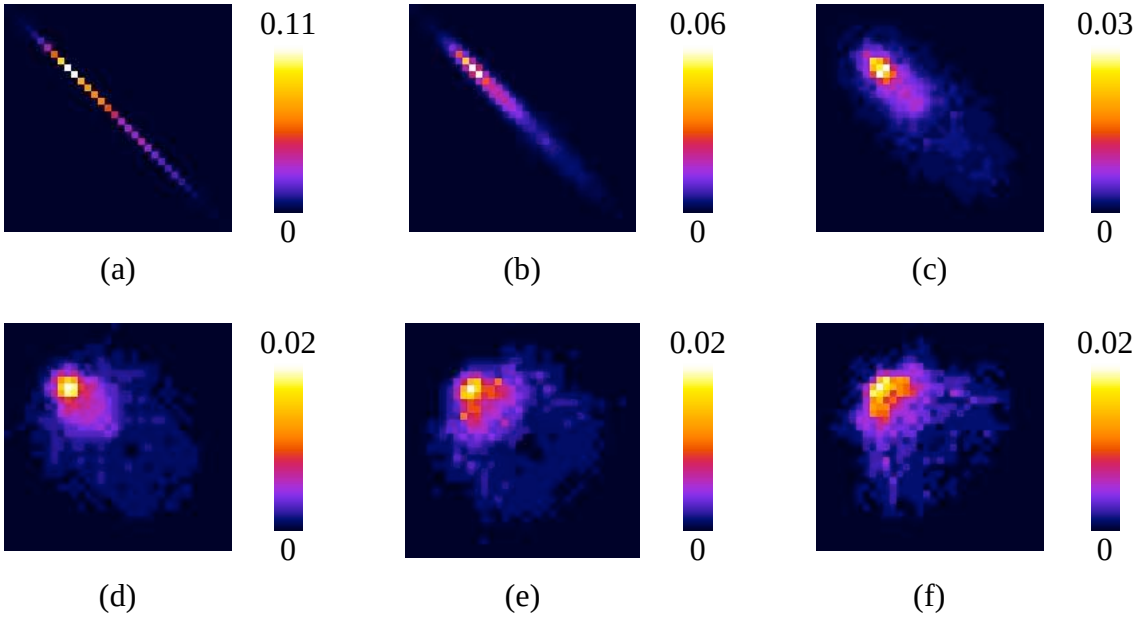



Figure 2-6 – Visual representation of GTSD matrices for horizontal nearest neighbours for the liver $R_2$ map in Figure 2-1.

The GTSD matrices were determined with pixel separations of (a) d = 0, (b) d = 1, (c) d = 5, (d) d = 10, (e) d = 15, (f) d = 20. The colour scale represents the probability of the co-occurrence of pixels with intensities given by the position in the matrix.

### 2.2.3 Texture Characterisation from GTSD Matrices

To use this technique for quantification purposes, the distribution of values in the GTSD matrices must be described numerically. A numerical description is achieved by applying a series of 2-D matrix algorithms that describe each GTSD by a single number. In the work of Haralick [137], fifteen measures are defined with each measure being sensitive to different aspects of texture in the image. Although the number of texture measures that can be derived from GTSD matrices is virtually unlimited, the

measures to be used in this research will be based on the fifteen measures defined by Haralick et al [137]. These measures will be referred to as texture parameters $t_1$ through to $t_{15}$ (the definitions of all fifteen measures are in the Appendix). As an example of these measures, consider texture parameter $t_1$,

$$t_1 = \sum_i \sum_j \left[ P(i,j)/R \right]^2$$

where $P(i,j)$ is the element of the GTSD matrix at row $i$ and column $j$ and $R$ is the total number of pixel pairs included in the GTSD matrix construction.

Texture measure $t_1$ is likely to be a measure of homogeneity or periodicity in any given image. The squared term effectively gives a higher "weighting" to those matrix elements that are larger due to the high number of occurrences of that intensity combination at that pixel separation.

## *2.3 Testing Texture Measures Derived from GTSD Matrices*

### 2.3.1 Test Patterns

To assess the efficacy of the various texture measures, a series of test images were created to give spatial variations of intensities similar to components of those seen in a typical liver $R_2$ map. These variations were achieved by generating images according to the equation

$$Intensity(x,y) = c + m\left[ \sin\left( \frac{2\pi x n}{w} \right) \sin\left( \frac{2\pi y n}{h} \right) \right]^i$$

where $c$ is the constant background intensity, $m$ is the magnitude of the periodic variations, $w$ and $h$ are the image width and height respectively, $x$ and $y$ are the co-ordinates of the pixel of interest in the image. The two remaining variables in this equation dictate the size and relative intensity of the variations in the image. The variable $n$ affects the frequency of the variations in the image while $i$ affects the effective radius of the peaks in the image.

Owing to the quantisation of the images prior to texture analysis, the same results were obtained regardless of the value assigned to the constant background intensity and the variation magnitude. For this reason, in all following analyses, the constant background and the magnitude of the periodic variations were both fixed at an intensity of 128. Additionally, as the typical size of an MRI image is 256 by 256 pixels, the height and width of the test images were both set to 256.

$$Intensity(x,y) = 128 + 128\left[\sin\left(\frac{2\pi xn}{256}\right)\sin\left(\frac{2\pi yn}{256}\right)\right]^{i}$$

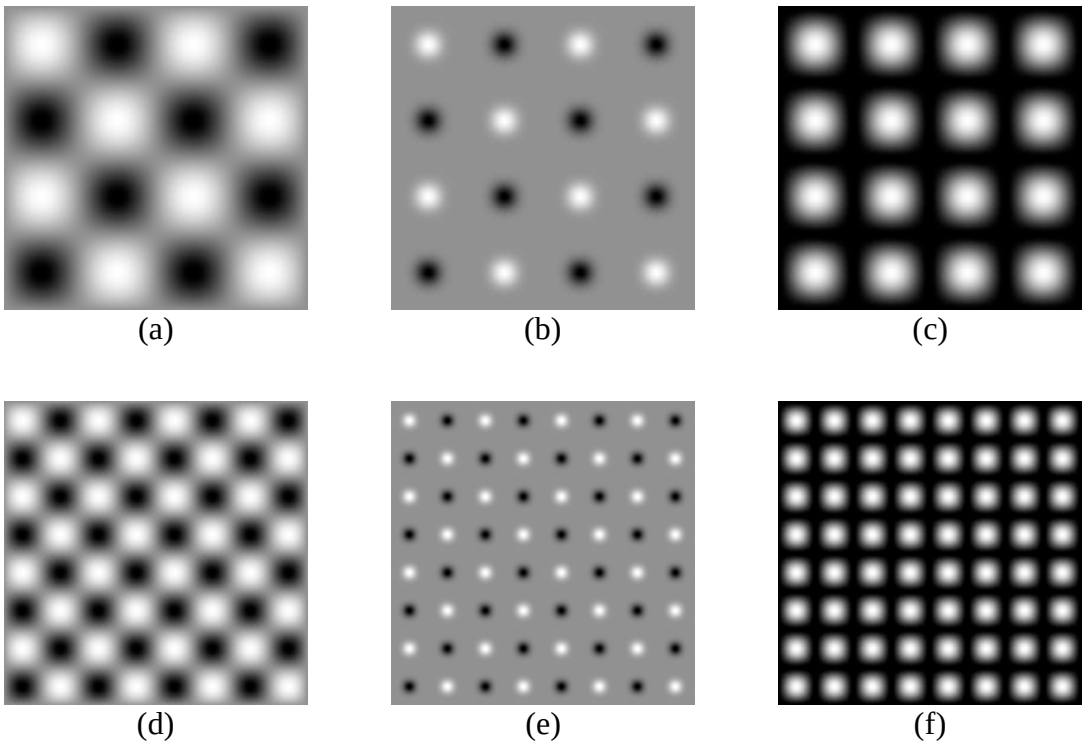


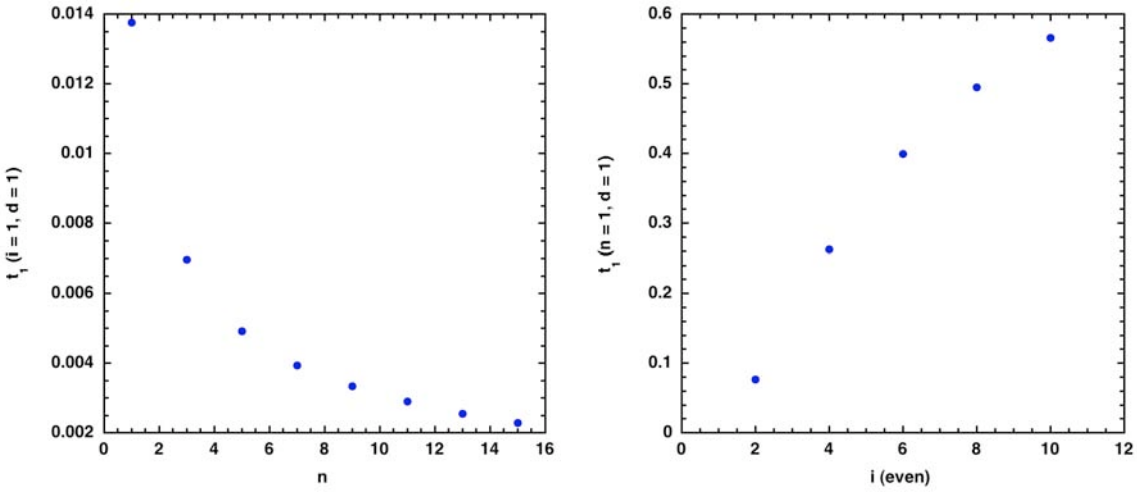

Figure 2-7 – Sample images on which the GTSD texture measures were assessed.

The parameters used to create the images were (a) $n = 2$, $i = 1$, (b) $n = 2$, $i = 7$, (c) $n = 2$, $i = 2$, (d) $n = 4$, $i = 1$, (e) $n = 4$, $i = 7$, (f) $n = 4$, $i = 2$.

The characteristics of the intensity variations in the image can be manipulated by changing the values of *i* and *n* in the equation. Changes in *n* create different numbers of intensity peaks in the test image, while increases in *i* cause the peaks to appear much sharper relative to the background. Odd values of *n* give images in which the intensity values can be above and below that of the background and even *n* images display only increases from the background intensity. As a result, the odd and even *n* are considered separately in the analyses to follow. Figure 2-7 illustrates the effects that varying *i* and *n* have in the images.

### 2.3.2 Results from Test Patterns

There are fifteen texture parameters that will be used in analyses in the following chapters, with each parameter calculated at a number of pixel separations. The tests in this section will only be carried out on texture parameters from GTSD matrices calculated using the nearest-neighbour pixels.




(a) (b)

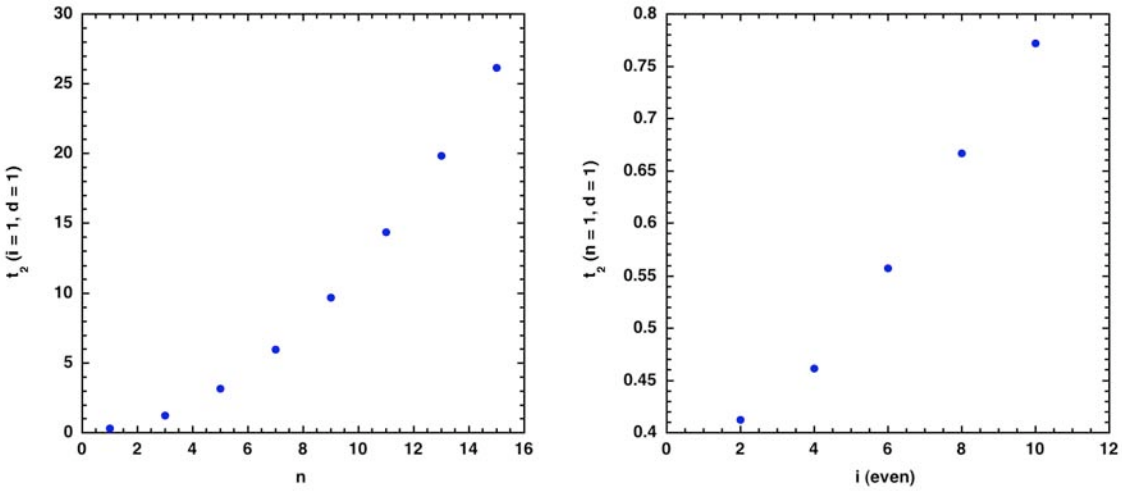
Figure 2-8 – Results from testing texture measure $t_1$ on sample images

(a) $t_1$ (derived from a distance 1 GTSD matrix) plotted against $n$ with $i = 1$, (b) $t_1$ (derived from a distance 1 GTSD matrix) plotted against even $i$ with $n = 1$.

As an initial example, consider texture parameter $t_1$ with the equation shown earlier. As the frequency of variation ($n$) increases, the value of $t_1$ tends to decrease with a trend similar to that of an exponential decay (Figure 2-8a). Additionally, as the relative size ($i$) of the variations increase, $t_1$ also increases (Figure 2-8b). The trend of increase in this case appears to be similar to that of a logarithmic function. From this example we can see that this parameter $t_1$ will be largest when the frequency of variations is low, and the relative size of the inhomogeneities is also small.

Texture parameter $t_2$, also at a pixel separation of 1 pixel increases in an exponential-like fashion as both $n$ and $i$ increase (Figure 2-9). So $t_2$ is most likely to be large when there is a high frequency variation in the image. This parameter will be accentuated further when the inhomogeneities are point like, as the radius decreases. These trends suggest that $t_2$ is weighted toward heterogeneity in an image. It appears that the effect of increasing frequency on $t_2$ is more dominant than the effect of effective radius. As $t_2$ increases more rapidly with $i$ as opposed to $n$, this dominance may not exist at large

values of *i*. However, in the context of $R_2$ images, the sharp edges introduced by large *i* are not expected, so the observations made here will be typical of that seen for the texture analysis of a liver $R_2$ map.



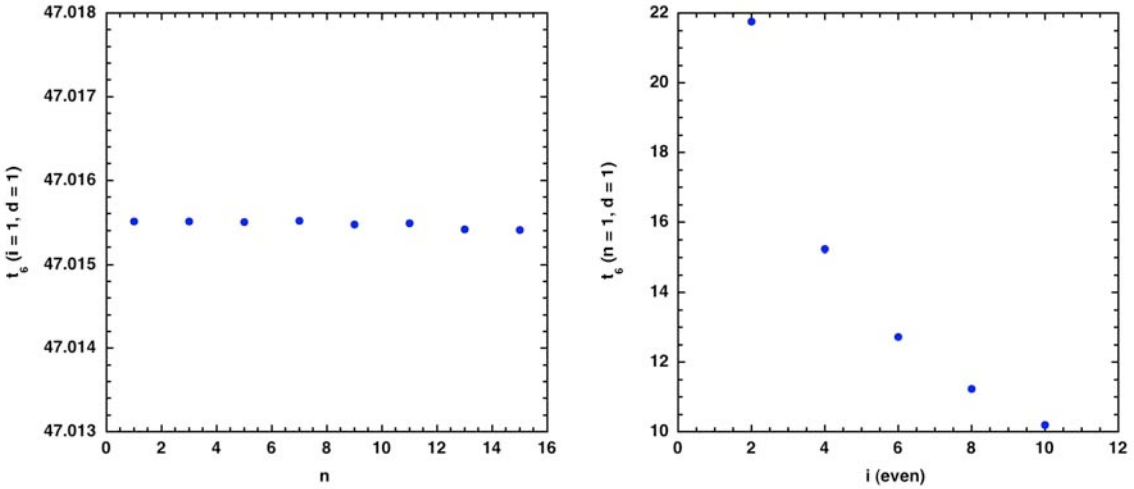

(a) (b)

Figure 2-9 – Results from testing texture measure $t_2$ on sample images

(a) $t_2$ (derived from a distance 1 GTSD matrix) plotted against *n* with $i = 1$, (b) $t_2$ (derived from a distance 1 GTSD matrix) plotted against even *i* with $n = 1$.

A final example is that of the texture parameter $t_6$ (Figure 2-10). As *n* increases, $t_6$ remains almost constant. In contrast though, a very rapid decay in $t_6$ is seen with increasing *i* as the effective radius of the variations decreases. So, one can assume that $t_6$ is most likely a measure of the degree of smoothness in an image as it is highest when there are fewer sharp edges.

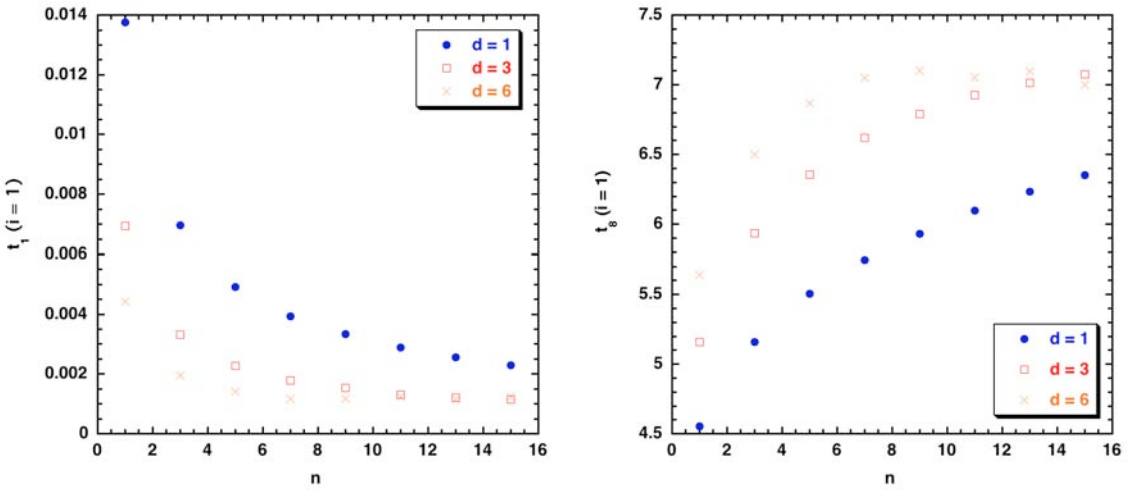
The effect on each texture parameter when considering pixel separations longer than just nearest neighbours is dependent on the individual parameter. While going to longer pixel separations causes some parameters to decrease (Figure 2-11a) others will increase (Figure 2-11b). However, the shape of the dependence on the frequency of variations *n* does not appear to differ as the pixel separation changes. In these test images, at larger pixel separations, the trend observed appears to flatten off at large *n*. This effect is most likely an artefact of the periodic nature of the test images and not attributable to any distinct texture feature.




(a) (b)

Figure 2-10 – Results from testing texture measure $t_6$ on sample images

(a) $t_6$ (derived from a distance 1 GTSD matrix) plotted against $n$ with $i$ = 1, (b) $t_6$ (derived from a distance 1 GTSD matrix) plotted against even $i$ with $n$ = 1.




(a) (b)

Figure 2-11 – Results from testing texture measures derived from GTSD matrices of varying distance on sample images

(a) $t_1$ plotted against $n$ with $i$ = 1 for distance 1 (●), 3 (□) and 6 (×) GTSD matrices, (b) $t_8$ plotted against $n$ with $i$ = 1 for distance 1 (●), 3 (□) and 6 (×) GTSD matrices.

## 2.4 Summary

Unlike ROI methods, $R_2$ mapping techniques provide information regarding the spatial variation of $R_2$ rather than just a mean value. From these maps, the variations and texture of the $R_2$ values in a given tissue can be quantified and compared with clinical and physiological factors. Grey-tone spatial-dependence matrices provide an appropriate measure for quantifying the texture in $R_2$ maps. In measurements on a series of test images, the texture measures derived from these matrices appear to be sensitive

to the variations seen in a typical $R_2$ map. Additionally, different texture measures are sensitive to particular texture features, providing a method to analyse various aspects of the $R_2$ distribution simultaneously.

# Chapter 3

# The Relationships Between Liver $R_2$ Image Texture and Liver Fibrosis

## 3.1 Introduction

In healthy liver tissue, iron present in the liver is stored in the parenchymal cells. In the development of fibrosis, iron deposition is observed in Kupffer cells, possibly a result of the phagocytosis of damaged, iron-containing red blood cells [145]. In the presence of cirrhosis, iron is accumulated within the regenerative nodules and not in the fibrous tissue [134]. So, at least on the cellular level, the presence of fibrous tissue can affect the spatial distribution of iron. Assuming that this distribution is fractal in nature, the spatial variation seen at the cellular level will also affect the structures at longer length scales. In liver relaxometry, variation of $R_2$ and hence iron has been seen on about the centimetre length scale but, apart from one case report [146], it has not been confirmed that this variation is the manifestation of the effects of fibrous tissue on iron deposits at the cellular level. The methods for quantifying texture defined and investigated in the previous chapter can be used to give insight into whether this is indeed the case.

Currently, the accepted method for monitoring fibrosis is by liver biopsy, which involves taking a core sample of liver tissue about 1/50 000 the size of the liver [24]. A histology section is then taken from this sample and stained. The histological section is analysed by a pathologist and subsequently given a grading to describe the level of fibrosis in the sample. There are two different criteria that are presently used for this staging, the Knodell criteria [28] and the METAVIR criteria [29]. Using the Knodell criteria, the pathologist assigns a score of 0 to 6 to the sample of interest with a score of ‘0’ indicating no fibrosis present and a score of ‘6’ indicative of advanced cirrhosis. With the METAVIR criteria, a specimen is categorised by one of five distinct levels (F0 to F4) with F0 indicating an absence of fibrosis and F4 indicating cirrhosis.

Regardless of the scoring system used, the evaluation of fibrosis by analysis of biopsy samples poses a number of problems. Firstly, the procedure of taking a needle biopsy is very uncomfortable for the patient and the procedure carries a degree of risk with up to

5% of patients experiencing complications and possibly requiring an extended stay in hospital [26]. The mortality rate for this procedure is between 1/1000 and 1/10 000 [24]. Secondly, in light of the discomfort and risk of the procedure, frequent assessments of fibrosis levels using this method are rare and not advisable. A method suitable for follow-up studies becomes important in clinical trials where it may be important to monitor progression of fibrosis at regular intervals. Thirdly, there is an added issue of sampling error as the score given to the biopsy sample may not be indicative of the entire liver and can lead to under-diagnosis of cirrhosis [23]. Finally, the methods of scoring are crude and highly subjective. Variability has been reported between fibrosis scores given by different pathologists when considering the same sample of liver [30]. To alleviate many of these issues, the ideal scenario will be a method of analysis that is both non-invasive and more quantitative rather than qualitative.

Radiographic techniques have been investigated previously as a potential candidate for monitoring the progression of liver fibrosis [64, 66], as the examination can be done non-invasively and no biopsy is required. If progression of fibrosis in the liver alters the iron distribution sufficiently on the macroscopic scale, then $R_2$-MRI should be appropriate for quantification of fibrosis, owing to the relationship between R2 and liver iron concentrations. The distribution of $R_2$ has been shown to relate to fibrosis staging via the coefficient of variation of $R_2$ values within the liver [97]. Texture measures calculated from grey-tone spatial dependence (GTSD) matrices can be considered as an extension of the histogram characterisation (Chapter 2) and may also prove useful in the quantification of fibrosis.

The work in this chapter looks at the relationship between texture parameters determined from the GTSD matrices of $R_2$ maps and liver fibrosis grading. This comparison will be addressed in iron-loaded patients, hepatitis C patients and normal controls. Also, a consideration of the factors that may compromise the effect that fibrosis has on the texture of the $R_2$ maps is presented.

## *3.2 Materials and Methods*

### 3.2.1 Subjects

To investigate the relationship between fibrosis score and $R_2$ texture, a cohort of 105 subjects (the test cohort) was recruited. This cohort was sub-divided into four groups of

subjects – haemochromatosis patients (n = 20, mean age = 49, age range, 17 - 74), thalassaemia patients (n = 46, mean age = 30, age range, 8 - 63), patients with hepatitis and no evidence of iron loading (n = 31, mean age = 42, age range, 21 - 59) and healthy controls (n = 8, mean age = 41, age range, 29 - 62). Of the thalassaemia patients, 23 had been splenectomised.

In this cohort, each of the subjects, except those in the healthy control group, had a liver biopsy to assess the staging of fibrosis. Each biopsy sample was assessed by a pathologist and graded according to the Knodell and METAVIR criteria. None of the healthy controls had a biopsy but they were all assumed to have a fibrosis staging of 0 in both criteria. The staging results were then used to assign each subject to one of two groups. The first group, Group 0, consists of all subjects assigned a Knodell score of zero or one. The second group, referred to as Group 1, includes all other subjects – those that were assigned a Knodell score of 2 or above. The population of each group is shown in Table 3-1. Subjects were also assigned to groups according to the METAVIR scale where scores of F0 and F1 are considered to be in the low fibrosis group and F2 – F4 are considered as having high fibrosis. Fibrosis staging using the METAVIR criteria was available for only 99 of the 105 subjects. The group allocations using the Knodell and METAVIR criteria were different in 10 of the 99 subjects. The results of this chapter will be based on the group allocations (Group 0 and Group 1) from the Knodell criteria.

| Patient Group | n (Total) | LIC (±SD) mg/g dw | LIC range | n (Group 0) | n (Group 1) |
|---|---|---|---|---|---|
| Haemochromatosis | 20 | 10.65 (7.44) | 2.11-23.58 | 10 | 10 |
| Thalassaemia | 46 | 18.24 (10.50) | 1.14-42.70 | 23 | 23 |
| Hepatitis | 31 | 1.13 (0.66) | 0.28-2.68 | 18 | 13 |
| Controls | 8 | 0.79 (0.37) | 0.55-1.22 | 8 | 0 |

Table 3-1 – High-grade and low-grade fibrosis allocations using the Knodell criteria and mean liver iron concentrations (LIC) for each of the patient sub-groups for the test cohort.

A second cohort of subjects (n = 10) was also used in this study (the reproducibility cohort). Each of the ten subjects in this cohort was scanned twice with each scan performed on consecutive days and on different scanners. None of these subjects had liver biopsy samples taken. Table 3-2 shows the number of subjects in each condition group.

| Subject type | n |
|---|---|
| Haemochromatosis | 2 |
| Thalassaemia | 5 |
| Thalassaemia Intermedia | 1 |
| Normal Controls | 2 |

Table 3-2 – Number of subjects in each sub-groups in the reproducibility cohort

### 3.2.2 MRI Image Acquisition and Analysis

All of the subjects recruited in this study were assessed for liver iron according to the method of Chapter 2. The scans were performed on various 1.5T scanners in a number of different centres. This method is highly reproducible on different scanners with experiments on phantoms giving a CV of 2.1% when performed on various scanners [97]. As this technique shows little variability between scanners it was not necessary to ensure scans were performed on the same scanner. Axial images of the liver were collected for each subject in both the test and reproducibility cohort with a repetition time of 2500ms and echo times of 6, 9, 12, 15 and 18ms. The slice thickness was 5mm and the minimum gap between slices was 5mm. The fields of view varied considerably between patients, as they were set according to the size of the subject.

Following acquisition, the images were analysed as described in Chapter 2 to obtain liver $R_2$ maps. The maps were generated by pixel-wise bi-exponential fitting as this method has previously been shown to give better fits in multi-compartment systems than mono-exponential fitting [141, 142]. For the test cohort, the liver $R_2$ map was obtained for only one slice of the liver – the slice with the largest cross-sectional area. However, for the reproducibility cohort, $R_2$ maps were generated for multiple slices of liver (7 – 9 slices) in each patient. This allowed inter slice variations to be taken into account when assessing the reliability of texture quantification techniques for monitoring fibrosis.

### 3.2.3 $R_2$ Map Texture Characterisation

The methods defined in the previous chapter were used to quantify the texture characteristics in each of the liver $R_2$ maps that were generated. To ensure that accurate comparisons can be made between subjects, the fields of view for each scan were normalised by linearly interpolating between neighbouring pixels prior to the GTSD

matrices being generated. Hence, all scans were analysed for texture with a field of view of 430mm giving an image resolution of 1.7mm per pixel.

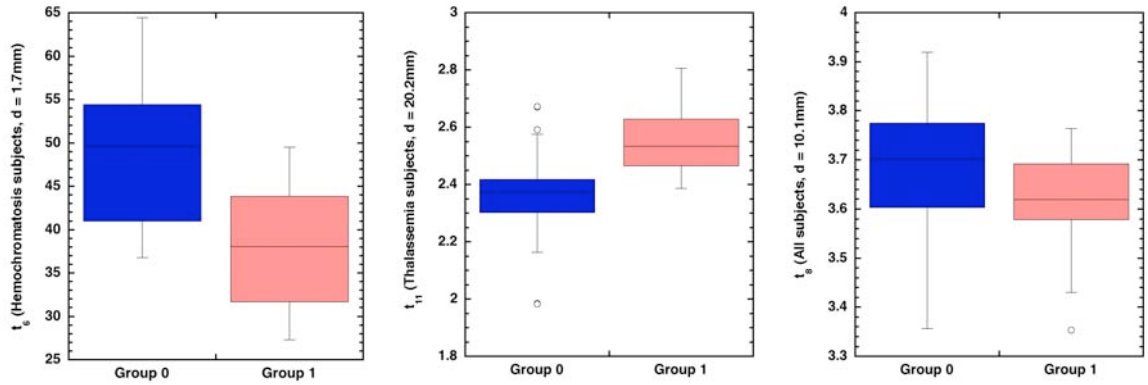
The GTSD matrices were generated for each patient at all pixel separations from 0 pixels (0mm) through to 20 pixels (33.6mm). Iron in the liver is not distributed homogeneously [21] but rather varies as a result of many factors including the presence of fibrous tissue. It is expected that the presence of fibrosis will contribute to the variations of $R_2$ on length scales up to a few centimetres, as the sizes of macroscopic nodules seen in advanced cirrhosis are on the centimetre length scale [147]. Exploring the effects at larger distances could provide more information on texture characteristics but doing this will decrease the region of interest in the texture analysis, potentially causing undesired effects.

## *3.3 Results*

### 3.3.1 Comparisons of Texture and Fibrosis for Different Patient Groups

Using the methods stated above, the effectiveness of each texture parameter at distinguishing between the two groups of patients was quantified. All texture parameters for GTSD matrices at all distances in the all subjects in the test cohort and also for the sub-groups of various conditions were assessed. As an example, the texture parameter $t_6$ at distance 1.7mm shows a distinction between group 1 and group 0 for the subset of haemochromatotic subjects (Figure 3-1a) but there is some overlap of the inter-quartile ranges (the boxes indicate the $25^{th}$ and $75^{th}$ percentiles). Wilcoxon's 2-sample test suggests that the distributions of $t_6$ for each group are significantly different ($p = 0.002$).

Another texture parameter, $t_{11}$ derived from a distance 12 (20.2mm) GTSD matrix shows no overlap of the inter-quartile range when considering only the subjects in the thalassaemia sub-group (Figure 3-1b). The distributions of $t_{11}$ for the two groups are significantly different ($p < 0.001$ from Wilcoxon's 2-sample test).




(a) (b) (c)

Figure 3-1 – Box-plots depicting the distinction between high-grade and low-grade fibrosis for texture parameters in different sub-groups of subjects

(a) $t_6$ (d = 1.7mm) for the haemochromatosis sub-group, (b) $t_{11}$ (d = 20.2mm) for the thalassaemia sub-group, (c) $t_8$ (d = 10.1mm) for the entire cohort. The boxes indicate the inter-quartile ranges and the error bars depict the 95% confidence intervals.

In looking for parameters that relate to fibrosis staging when all patients in the test cohort are combined, not many are successful. Of the parameters that show a statistically significant result, there is still considerable overlap between the two groups. As an example, the parameter $t_8$, at distance 6 (10.1mm) gives a substantial overlap of the inter-quartile ranges of the two groups (Figure 3-1c). Although the distributions of $t_8$ are statistically different (p = 0.02 from Wilcoxon's 2-sample test), the overlap of the inter-quartile ranges means that a parameter such as this is probably not clinically useful.

The discriminating power for fibrosis was assessed in all of all of the texture measures. Figure 3-2 shows the p values for every texture measure observed in the three disease conditions and also for the entire test cohort. All the p values are calculated from Wilcoxon 2-sample test statistic, testing for significant differences in the distributions of the texture parameters of the high and low fibrosis groups. Significant p-values are depicted in the yellow and red part of the scale, while insignificant p-values are shown in blue.

The p-values for the thalassaemic subjects in the cohort are shown in Figure 3-2a. Most of the texture parameters show highly significant differences in the distributions for the high-grade and low-grade fibrosis groups. In the case of the haemochromatotic subjects (Figure 3-2b), there appear to be a few parameters displaying some statistical

significance but not nearly as many as the thalassaemia group. Figure 3-2c shows the p-values for the hepatitis group of patients. It shows that there is no statistically significant separation between the groups for any of the parameters considered in this work.

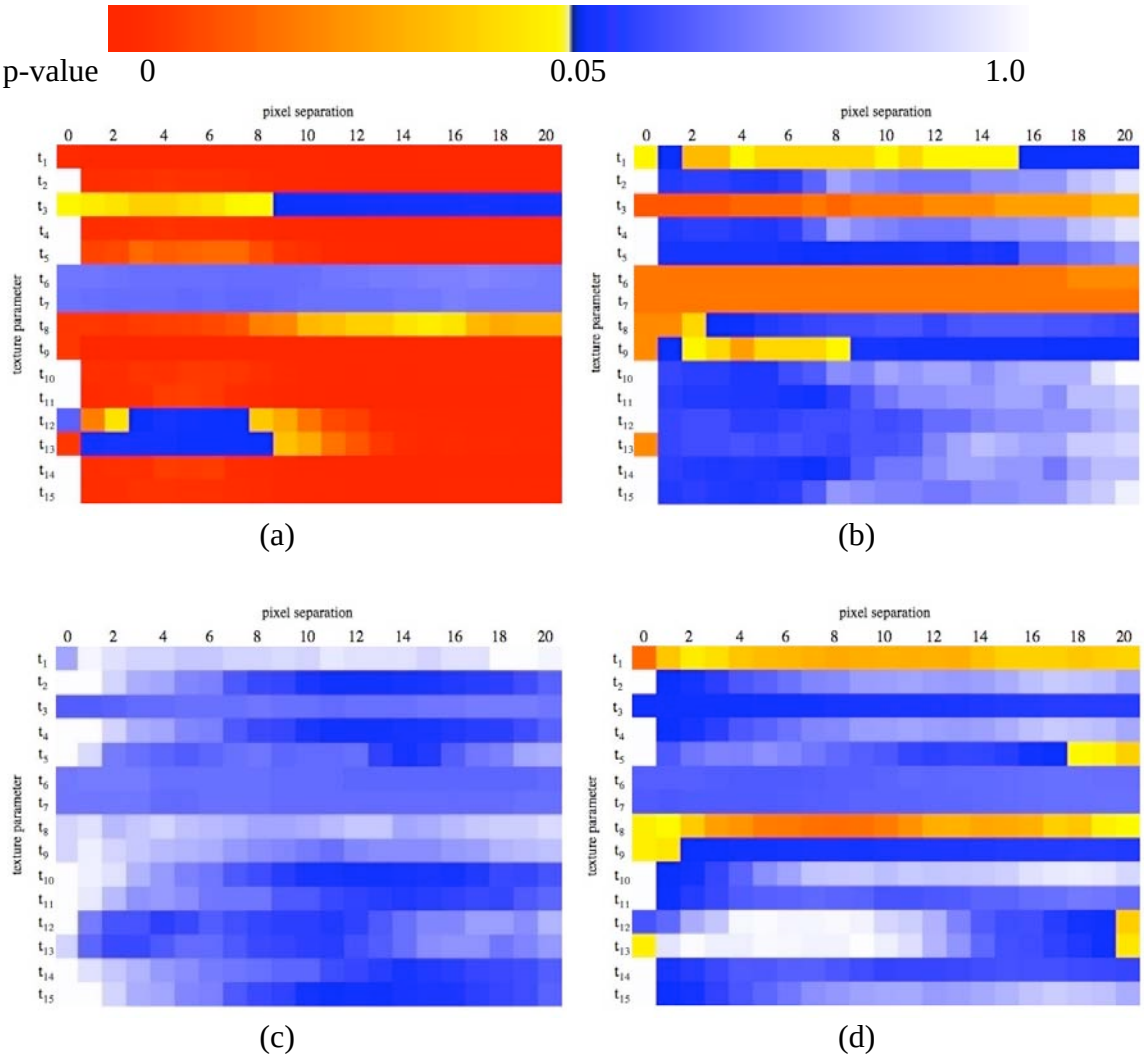



Figure 3-2 – The results of Wilcoxon's 2-sample test for all texture parameters in each sub-group

(a) the thalassaemia sub-group, (b) the haemochromatosis sub-group, (c) the hepatitis sub-group, (d) the whole test cohort. The colour scale depicts the probability (p) that the distribution of the texture parameter in group 0 is not significantly different from the distribution of the texture parameter in group 1. The distributions are considered to be significantly different for $p < 0.05$. A red-yellow colour scale is used for $p < 0.05$ and a blue-white scale is used for $p > 0.05$.

The information depicted in Figure 3-2d shows the statistical significance of a difference in the distributions of each texture parameter for Group 1 and Group 0 when

all the subjects, including the controls, are looked at as a whole. In this case, only two parameters, $t_1$ and $t_8$, show p-values less than 0.05 for GTSD matrices of all distances.

### 3.3.2 Independence of Texture Measures

#### Correlations Between Parameters

Although each of the texture parameters uses different equations to describe the GTSD matrices, there are some cases where one texture parameter is sensitive to similar (or the same) features as another. Figure 3-3 shows parameter $t_7$ versus $t_6$ with the points representing calculated texture values derived from GTSD matrices of all distances for every subject of the test cohort. Spearman's Rho test shows that these two parameters are highly and significantly correlated ($r = 0.999$, $p < 0.001$).

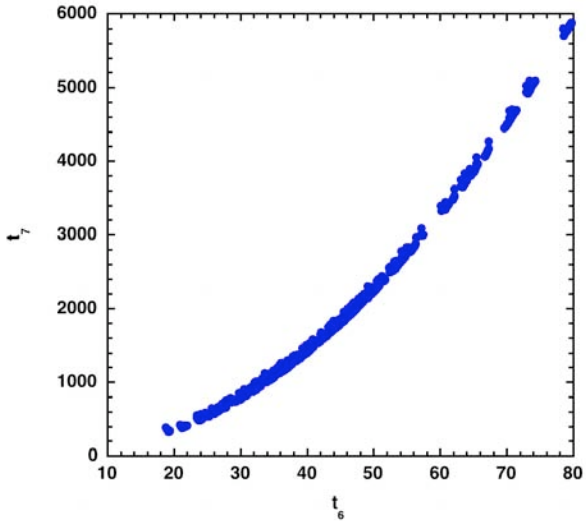



Figure 3-3 – Texture parameter $t_7$ plotted against texture parameter $t_6$.

Derived from GTSD matrices of all distance from 1 – 20 pixels, for all subjects in the test cohort.

#### Correlations Between Texture Parameters with Different Pixel Separations

Figure 3-4 shows how the correlation coefficient, r, varies for each parameter when compared to the corresponding parameter calculated from the distance 1 GTSD matrix. While some parameters appear to vary little as the pixel separation changes (for example $t_3$, $t_6$ and $t_7$) others such as $t_{10}$ show a much larger reduction in correlation only a few pixels away.

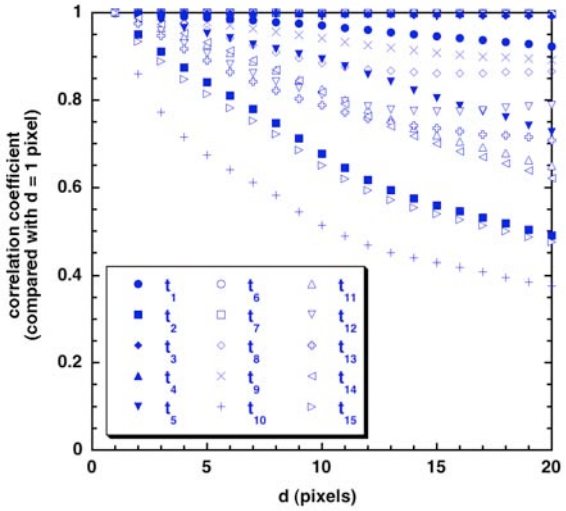
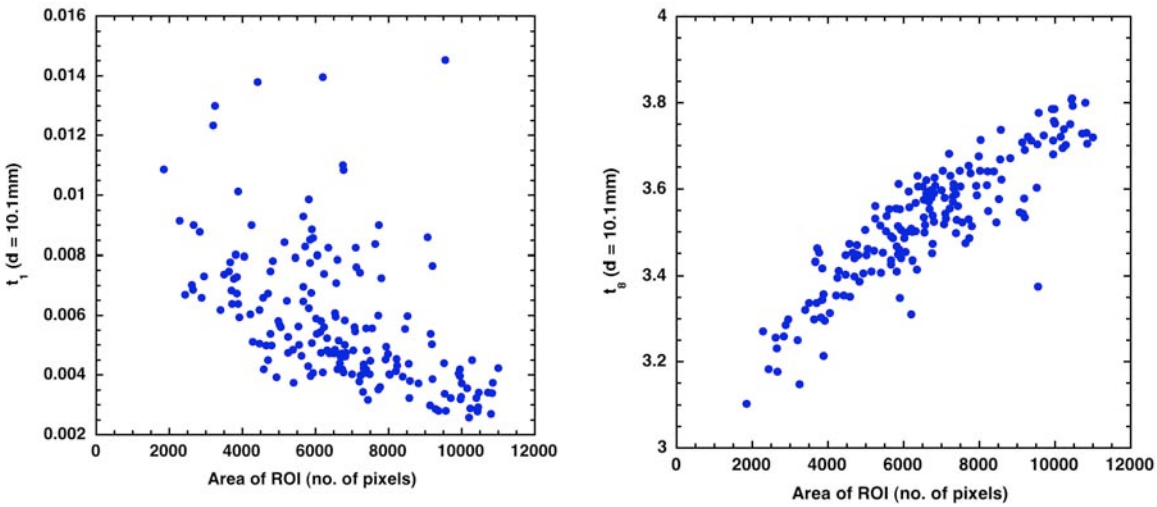



Figure 3-4 – Correlation coefficient of texture parameters $t_n$ (d = 1) with $t_n$ at all other distances from 2 to 20.

### 3.3.3 Effect of Region of Interest Area on Texture Measures

To investigate effects owing to the size of the region of interest, the results from the reproducibility cohort were analysed. As the analyses for these patients involved calculating the $R_2$ maps for numerous slices in the same liver, the inter-slice effects on the texture characterisation methods can be assessed.




(a) (b)

Figure 3-5 – Texture parameters (a) $t_1$, (b) $t_8$, plotted against the number of pixels in the region of interest.

Both parameters derived from distance 6 (10.1mm) GTSD matrices. All individual slices analysed in the reproducibility cohort are included in each plot.

Figure 3-5a shows how one parameter, $t_1$ from a distance 6 (10.1mm) GTSD matrix, varies with the number of pixels in the $R_2$ map. In this case, as the area of the slice

increases, this parameter generally tends to decrease. For some parameters however (Figure 3-5b), the parameter of interest (for this example $t_8$ distance 6) increases as the area of the region of interest increases. In this case, there is less scatter indicating that this parameter is highly sensitive to variations in area. As with the previous example, it appears that parameter $t_8$ also seems to become less affected by area changes as the area increases, although not to the same extent as $t_1$.

Some parameters, unlike the previous examples, show little or no variation caused by variations in the area of the region of interest. Figure 3-6 shows how the parameter $t_{12}$ (distance 6) changes with varying region of interest area. Although there is a lot of scatter in the plot, there does not appear to be any relationship with area.

Table 3-3 shows how the relationship between each texture parameter and the area of the ROI changes as the smaller slices are removed. When all slices are considered, regardless of ROI area, most parameters correlate with pixel area with high statistical significance. Considering only areas with 8000 pixels or more removes the dependence on area for all but three of the texture parameters.

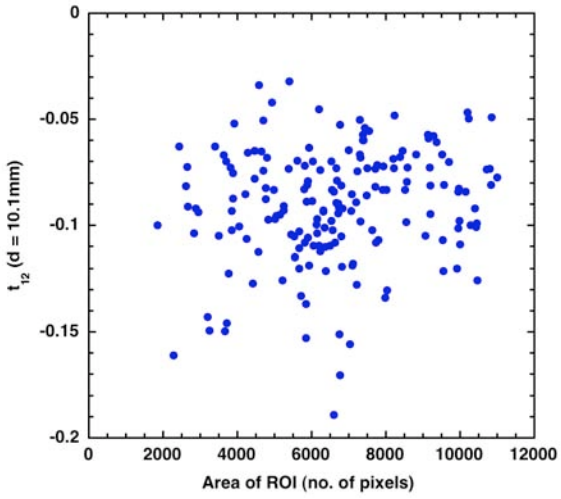



Figure 3-6 – Texture parameters $t_{12}$ plotted against the number of pixels in the region of interest.

Both parameters derived from distance 6 (10.1mm) GTSD matrices. All individual slices analysed in the reproducibility cohort are included in each plot.

| Texture measure | All slices | | Area > 6000 | | Area > 7000 | | Area > 8000 | |
|---|---|---|---|---|---|---|---|---|
| | ρ | p | ρ | p | ρ | p | ρ | p |
| $t_1$ | *-0.64* | *<.0001* | *-0.60* | *<.0001* | *-0.52* | *<.0001* | *-0.46* | *0.0017* |
| $t_2$ | *0.69* | *<.0001* | *0.52* | *<.0001* | 0.19 | 0.11 | -0.04 | 0.77 |
| $t_3$ | *0.58* | *<.0001* | *0.41* | *<.0001* | *0.50* | *<.0001* | *0.33* | *0.026* |
| $t_4$ | *0.69* | *<.0001* | *0.52* | *<.0001* | 0.19 | 0.11 | -0.04 | 0.77 |
| $t_5$ | *-0.50* | *<.0001* | *-0.48* | *<.0001* | *-0.34* | *0.0035* | -0.23 | 0.13 |
| $t_6$ | *0.37* | *<.0001* | *0.24* | *0.011* | *0.39* | *0.0006* | 0.17 | 0.28 |
| $t_7$ | *0.39* | *<.0001* | *0.26* | *0.0073* | *0.40* | *0.0004* | 0.18 | 0.25 |
| $t_8$ | *0.87* | *<.0001* | *0.69* | *<.0001* | *0.69* | *<.0001* | *0.65* | *<.0001* |
| $t_9$ | *0.80* | *<.0001* | *0.71* | *<.0001* | *0.65* | *<.0001* | *0.52* | *0.0003* |
| $t_{10}$ | *0.69* | *<.0001* | *0.47* | *<.0001* | 0.12 | 0.33 | -0.12 | 0.45 |
| $t_{11}$ | *0.66* | *<.0001* | *0.54* | *<.0001* | *0.29* | *0.012* | 0.10 | 0.50 |
| $t_{12}$ | *0.15* | *0.046* | *0.23* | *0.015* | -0.01 | 0.93 | -0.16 | 0.30 |
| $t_{13}$ | -0.02 | 0.82 | -0.15 | 0.11 | 0.07 | 0.56 | 0.20 | 0.19 |
| $t_{14}$ | *0.65* | *<.0001* | *0.54* | *<.0001* | *0.30* | *0.0093* | 0.10 | 0.51 |
| $t_{15}$ | *0.69* | *<.0001* | *0.51* | *<.0001* | 0.19 | 0.12 | -0.05 | 0.76 |

Table 3-3 – Results from Spearman's rank order test for correlations between area of the region of interest and texture parameters.

All parameters considered were derived from distance 6 (10.1mm) GTSD matrices. Results in italics identify statistically significant correlations.

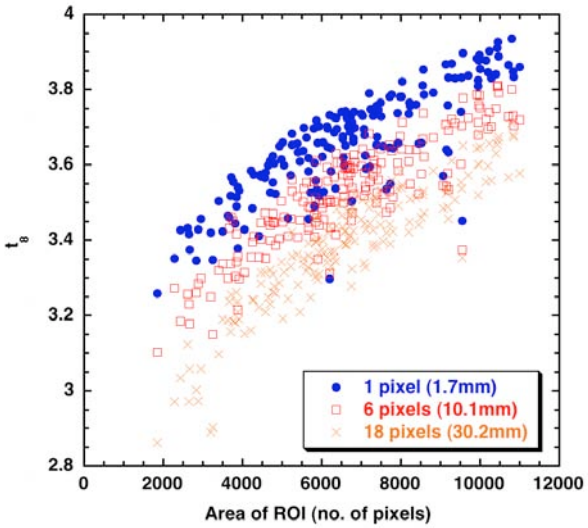



Figure 3-7 – Texture parameter $t_8$ plotted against area of the region of interest for various pixel separations.

Pixel separations of 1 (●), 6 (□) and 18 (×) pixels are shown. Texture parameters were measured in all individual slices analysed in the reproducibility cohort.

Table 3-3 shows the parameters at only one distance (6 pixels). However, a look at the relationships seen at other distances shows very similar trends to those shown here. An example of this is shown in Figure 3-7, where texture parameter $t_8$ was plotted for distance of 1, 6 and 18 pixels (1.7mm, 10.1mm and 30.2mm respectively). Although the

value of the parameter changes at different distances, the general increasing trend is common to all three plots. This relationship with distance seen here is typical of all other texture parameters.

### 3.3.4 Inter-patient and Intra-patient Variations of Texture Measures

The results from the reproducibility cohort were used to compare the variations of the texture parameters seen within a single patient to the variations seen between the patients. In these results, the two scans carried out on each patient are considered as independent events. Each subject has a large number of texture parameters associated with it. Each of the texture measures can be represented as $t_n(d, s)$ where $t_n$ is the parameter of interest, $d$ is the pixel separation used to derive the GTSD matrix and $s$ is the slice for which the parameter is calculated. As each of the slices in a scan are of the same liver and at the same time $t_n(d)$ can be defined as the value of the parameter at a specified pixel separation averaged over all slices in the scan. By calculating the coefficient of variation for each patient ($CV_p$), the variation seen over all slices within a single scan can be quantified. The mean $CV_p$ over all patients gives an indication of the variation generally expected between different slices of a single scan. This value will be referred to as the inter-slice coefficient of variation ($CV_{is}$).

The inter-patient coefficient of variation ($CV_{ip}$) describes how a particular parameter will vary between patients after variations due to inter-slice effects have been averaged out. $CV_{ip}$ is calculated by finding the coefficient of variation on $t_n(d)$ over all patients.

Figure 3-8 shows the difference between the $CV_{ip}$ and $CV_{is}$. In this figure a positive result indicates that the inter-patient effects are dominant. The results from the graph show that most parameters are dominated by the variation between slices. Parameters $t_3$, $t_6$ and $t_7$ are more affected by inter-patient effects for all pixel separations. However, parameters such as $t_1$ and $t_{12}$ show much larger dependence on inter-slice effects especially for larger pixel separations with $CV_{is}$ over 15% greater than $CV_{ip}$ in some cases. In fact, all of the parameters become more affected by inter-slice variations as the pixel separations increase.

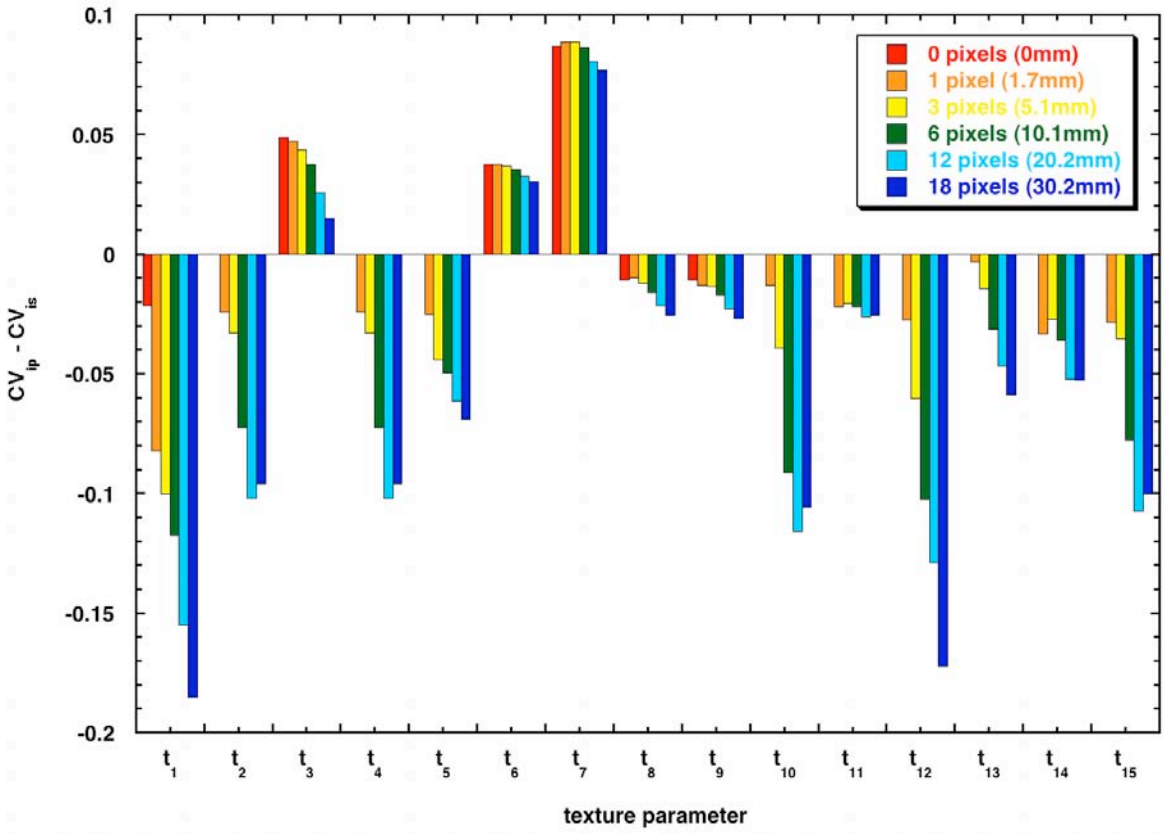



Figure 3-8 – Comparison of inter-patient and inter-slice effects on texture measures.

The difference between the mean inter-patient co-efficient of variation ($CV_{ip}$) and the mean inter-slice co-efficient of variation ($CV_{is}$) in the reproducibility cohort for texture parameters derived from GTSD matrices of selected distances.

For an accurate comparison of inter-patient and inter-slice effects, the dependence on slice area should be minimised in such an analysis. The comparison of $CV_{is}$ and $CV_{ip}$ was repeated with the CV's calculated using only the four largest slices of each scan rather than all of the slices. In using the largest slices, it is assumed that the slice areas are such that all parameters are mostly invariant with slice area changes. The results are shown in Figure 3-9.

After removing the effects of small slice areas, it has become apparent that the majority of parameters are dominated by the inter-patient variations. The parameters $t_5$, $t_{13}$ and $t_{14}$ are notable exceptions, especially at the higher pixel separations.

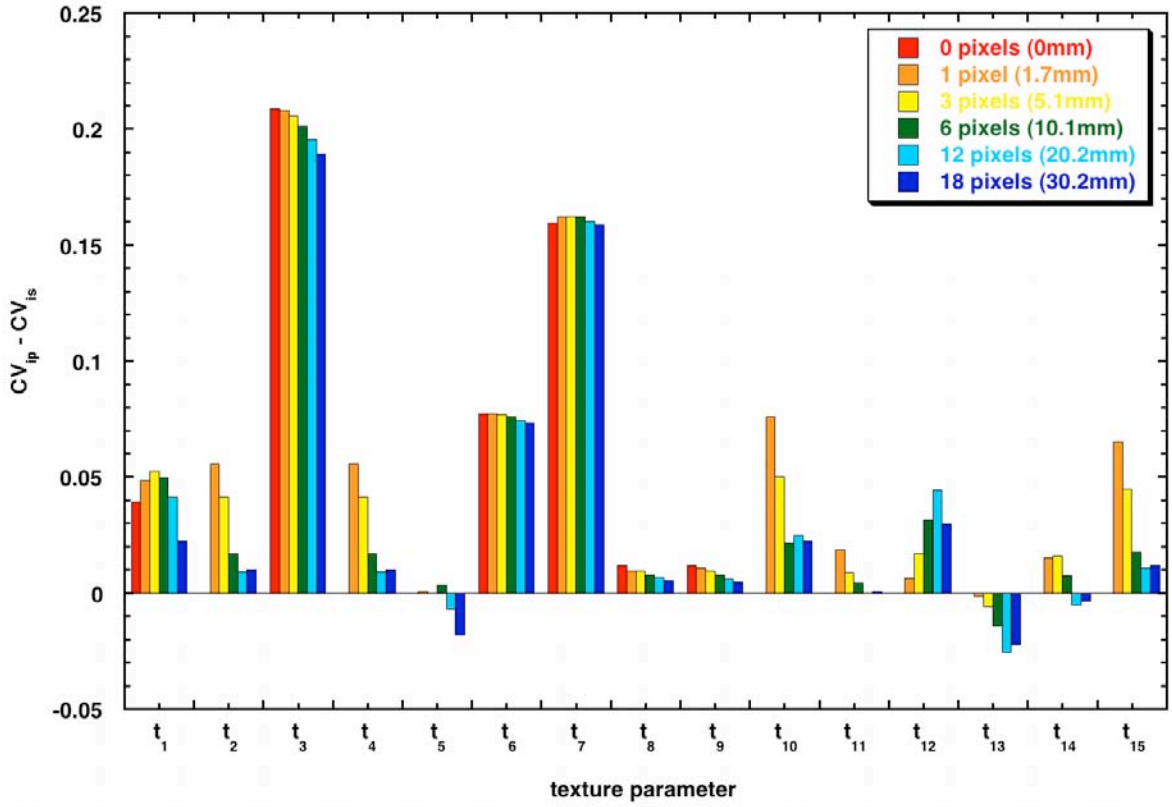



Figure 3-9 – Comparison of inter-patient and inter-slice effects on texture measures using only large regions of interest.

The difference between the mean inter-patient co-efficient of variation ($CV_{ip}$) and the mean inter-slice co-efficient of variation ($CV_{is}$) in the reproducibility cohort for texture parameters derived from GTSD matrices of selected distances, considering only the four largest slices from each scan.

## *3.4 Discussion*

Results from both the Knodell and METAVIR methods were obtained for the patients in this study. The assignment of the patients to a high or low fibrosis group does depend on the criteria used to grade the samples. However, disagreements in the assignment of the patients to these two groups using the two different criteria were observed in less than 10% of the patients. Although this disagreement will introduce variations into the results depicted in this study, the general observations are likely to be unchanged.

### 3.4.1 Effect of Iron-loading on Texture Measures

Some of the texture parameters show a statistically significant distinction between the two groups of fibrosis staging while others exhibit no distinction at all. Also, a given texture parameter that could potentially distinguish between the high and low fibrosis cases in one of the patient sub-groups shows a poor relationship with fibrosis in another

sub-group. In the case of the sub-groups of patients with iron overload, it appears that in general, the parameters that exhibit a statistically significant difference between the high and low fibrosis patients in the thalassaemia group do not distinguish between the fibrosis grade in the case of the haemochromatosis group and vice versa. Iron loading in haemochromatosis patients is a result of unregulated dietary iron absorption while in the thalassaemias iron is accumulated owing to repeated blood transfusions or increased erythropoiesis [13, 148]. These different iron-loading mechanisms can cause different distributions of iron in the liver tissue at the microscopic length scale, possibly giving a variation in the texture of the $R_2$ maps. The texture changes owing to development of fibrosis in the liver tissue will differ according to the underlying iron-loading mechanism. So the parameters used to monitor fibrosis are likely to depend on the underlying pathology of the patient.

The results suggest that these methods are more sensitive to fibrosis staging in the presence of iron overload conditions. A large number of texture measures give rise to statistically significant distinction between the fibrotic and non-fibrotic groups in the case of the haemochromatosis and thalassaemia sub-groups. In contrast, this is true for only a small number of texture measures in the hepatitis sub-group where the average iron concentrations are typically lower. The quantisation of the images prior to generating the GTSD matrices also acted to normalise the intensities in the $R_2$ maps, so any enhanced sensitivity due to the inherently high $R_2$ values in the iron-loaded sub-groups was eliminated. It can be assumed then that the iron in the liver effectively acts as a natural contrast agent for fibrosis. The presence of fibrous tissue will result in a decrease in the local iron concentration and hence a hypo-intense $R_2$ value. This effect will be more obvious in the presence of high iron concentrations as the intensity differences in the $R_2$ map will be greater and will not be removed by the normalisation process.

Many more parameters were observed to distinguish between the two groups with sufficient statistical significance in iron-loaded sub-groups as compared to the whole patient population. The smaller number of texture measures appropriate for distinguishing fibrosis staging in the whole population is most likely a result of the effects of the patients with low iron concentrations masking the effects seen in those with high iron concentration. The high proportion of iron-loaded patients in this study (63%) contributes most to the entire patient population. This high proportion coupled

with the increased sensitivity of some measures to fibrosis staging in the presence of iron in both iron-loading mechanisms could be responsible for the apparent power for distinguishing between the two fibrosis groups in the entire patient population. However, there are a number of parameters that appear to be indicative of fibrosis staging in only one of the iron-loaded sub groups. As the haemochromatosis and thalassaemia groups only account for 19% and 44% of the whole study population respectively, the effect seen for a parameter in only one of the iron-loaded groups may be diluted in the whole population and hence not seen as an appropriate parameter. This effect is further evidence that the iron may act as a contrast agent for monitoring fibrosis using MRI texture based techniques.

### 3.4.2 Factors Other Than Fibrosis

Some parameters show a distinct correlation with each other regardless of the patient in which they were measured, the condition of the patient or the pixel separation used to determine the GTSD matrix. In the cases where this type of relationship exists, it can be assumed that the two parameters are sensitive to the same texture feature and cannot be considered as independent texture measures. Additionally, texture parameters measured for similar pixel spacings are also closely related in all of the parameters, owing to the smooth variations seen in the $R_2$ maps. These similarities mean that caution must be taken in assuming that different pixel separations used in measuring a single parameter are independent measurements. The differences in the measurements will depend on which parameter is being measured and how different the two pixel separations are. Identifying redundancies resulting from correlations between texture parameters is important if these parameters are to be combined to increase the sensitivity of texture characterisation to fibrosis staging. Caution should also be exercised in the interpretation of the statistics in comparisons of multiple parameters. The quoted values for statistical significance, while illustrating that iron distributions as measured from R2 maps provide information regarding the level of fibrosis, are not useful in evaluating the utility of an individual parameter. The definitions of statistical significance would have to be altered to take into account the higher probability of a favourable outcome for a single variable when so many are considered at once.

The use of GTSD matrices to measure texture allows the whole slice of the $R_2$ map to be used in the analysis regardless of size or geometry. However, the effective region of interest used in calculating the GTSD matrices will decrease as the pixel separation

used to determine the GTSD matrices increases. As pixel separation increases, more pixels at the boundary of the $R_2$ map are excluded from the analysis potentially introducing some sampling error, as the whole slice is no longer considered. This effect would be even more severe in the case where the number of pixels that make up the $R_2$ map is small. So in texture measures calculated from GTSD matrices derived from $R_2$ maps with a small number of pixels, caution must be taken in interpreting the values, especially when the GTSD matrix is derived form a large pixel separation. The sampling deficiency in the smaller slices is one possible cause of the variations seen in the texture parameters when the ROI used to calculate it only has a small number of pixels.

The sensitivity of GTSD-derived texture measures for fibrosis may also be limited by the image acquisition methods employed. While the method used is sufficient for the quantification of mean liver iron concentrations in the liver [97], it is not ideal for assessing fibrosis by texture analysis. Most notably, the 5x5 smoothing kernel used in this method lowers the effective resolution of the image and averages out the variations that are postulated to occur as a result of the presence of fibrosis. This smoothing will manifest in the GTSD matrices by weighting the probabilities toward the main diagonal of the matrix. In turn, the texture measures will also have values associated with a more diagonal matrix than is expected, lowering the sensitivity. Any further investigation into these techniques should address this issue and test these texture analyses on more appropriate image acquisition methodologies.

Variations in area can affect the value of a given texture parameter. The majority of the parameters considered in this study are sensitive to area variations at all pixel separations. Although, the extent to which these variations affect the texture parameter depends on both the parameter of interest and the pixel separation used to calculate the texture measure. The parameters that show a distinct relationship with area will not be suitable for grading fibrosis, as the variations in texture are unlikely to result from effects due to fibrosis. On the other hand, variations in texture parameters that are unaffected by area are far more likely to be due to fibrosis. Parameters such as these could prove to be useful in comparisons with fibrosis staging as the choice of the slice used for analysis will not affect the value of the texture parameter or mask any relationship that may exist with fibrosis staging. In fact, most parameters are free from complications related to ROI area provided the ROI is chosen so that it includes over

8000 pixels. Ensuring that a slice is chosen such that the area of the ROI is sufficiently large will increase a given texture parameter's sensitivity to fibrosis.

The grading of liver fibrosis from biopsy-sized samples is not homogeneous throughout a liver and has been found to vary considerably [23]. It is unlikely that sampling error effects will be as large when comparing slices of liver tissue rather than biopsy-size samples. However, the variations seen in the texture parameters are mostly dominated by the intra-patient variations, attributable to the texture parameters being affected by slice area. The removal of slices of smaller area results in an increase in sensitivity to the variations between subjects but the variation due to the intra-patient effects are still present and significant. As the majority of the subjects in the reproducibility cohort are not iron loaded, the dominance of intra-patient variations may result from the low sensitivity of the texture parameters to fibrosis in the absence of iron. Repeating the same analyses on a sufficiently large cohort of only iron-loaded subjects may see improved results owing to the enhancement of the fibrosis-texture relationship in the presence of iron loading.

## 3.5 Summary

Texture quantification of $R_2$ maps using GTSD matrices is a potential method to non-invasively monitor the progression of fibrosis in the liver. Some of the texture measures derived from this method appear to distinguish between patients with high and low grade fibrosis. Variations in texture measures due to factors other than fibrosis and the lack of sensitivity in the absence of iron limit the clinical utility of these methods. While no individual parameter is likely to be clinically useful in the grading of fibrosis, a combination of parameters in a model may enable a test with clinical utility. An attempt at developing such a model is derived in the next chapter.

# Chapter 4

# Development of a Model For the Prediction of Liver Fibrosis Using Liver $R_2$ Image Texture

## 4.1 Introduction

The previous chapter addressed how the existence of fibrosis can affect various aspects of texture in a liver $R_2$ map. Some of the texture parameters investigated did appear to distinguish patients with little or no fibrosis from those showing evidence of the presence of fibrosis. When the test cohort was split into groups by disease type, it was found that the power of texture parameters for predicting the presence of fibrosis was increased in the patients suffering from iron loading conditions. However, it was also revealed that factors other than fibrosis can also affect the texture measures and may confound any relationship between texture and fibrosis.

Rather than using individual texture measures as indicators of fibrosis, combinations of the individual texture parameters may enhance the predictive power of texture measures for fibrosis. Models developed from combinations of serum tests have proven to be indicative of fibrosis staging [42, 49, 50] and a similar approach may also be of use in texture characterisation.

In this chapter a statistical method, will be employed to generate a candidate model for the prediction of fibrosis using a combination of the texture parameters. The reliability, as well as the limitations, of monitoring fibrosis staging through texture characteristics will also be addressed.

## 4.2 Materials and Methods

### 4.2.1 Subjects

The subjects used in this study are the same two cohorts of patients used in the previous chapter. The test cohort will be used as a training set to develop a model for predicting fibrosis using logistic regression methods. The reproducibility cohort will be used to test the reproducibility of these measurements between two examinations, performed on

different scanners on consecutive days. Although slice area effects proved to be a complicating factor in texture measurements derived from grey-tone spatial dependence (GTSD) matrices, all subjects in the test cohort were analysed using the slice with the largest cross-sectional area. It is assumed that this will greatly minimise the slice area effects discussed in the previous chapter. In contrast, for the reproducibility cohort, multiple slices were investigated in each scan.

As in the previous chapter, all subjects in the test cohort were assigned to either a high or low fibrosis group based on grading of biopsy specimens using the Knodell and METAVIR fibrosis staging criteria. Less than 10% of the patients in the cohort were allocated differently by the two methods and the differences in the statistical comparisons of the data using each of the two criteria were negligible. Owing to the minimal discrepancy between the two criteria and the fact that many non-invasive tests for fibrosis have been verified with the METAVIR system, in the results following, only the results obtained using the METAVIR criteria will be presented.

### 4.2.2 Texture Analysis

To quantify the texture of the $R_2$ maps generated for each patient, the methods outlined in Chapter 2 and employed in the analyses of Chapter 3 will be used here. These texture measures will then be used in the logistic regression methods to generate a candidate model for fibrosis prediction.

### 4.2.3 Statistical Methods

In this study, a logistic regression model was used to develop a parameter with a higher predictive power for fibrosis than any single texture measure. This model was found to be most appropriate for these data as other regression models are not appropriate for use in the prediction of a dichotomous variable.

While a brief summary of the principles of logistic regression will be discussed, a more detailed account of logistic regression theory and its applications to clinical practice can be found elsewhere [149].

The logistic regression model is depicted mathematically by the equation

$$\pi(\tilde{x}) = \pi(x_1, x_2, ..., x_n) = \frac{e^{\beta_0 + \beta_1 x + ... + \beta_n x}}{1 + e^{\beta_0 + \beta_1 x + ... + \beta_n x}}$$

A transformation of this equation known as the logit transformation is useful to the methods of logistic regression and is given by

$$g(\tilde{x}) = g(x_1, x_2, ..., x_n) = \ln\left[\frac{\pi(\tilde{x})}{1 - \pi(\tilde{x})}\right]$$

where $g(\tilde{x}) = \beta_0 + \beta_1 x + ... + \beta_n x$ and $n$ is the number of variables

The logit function $g(\tilde{x})$ is useful in light of the fact that unlike $\pi(\tilde{x})$, it is a linear function and hence has similar properties to that of linear regression.

As with any statistical method, assumptions need to be made about the properties of the sample of interest. The main concern in a sample like that used in this study is the issue of sample size. If the 99 subjects of the test cohort for which METAVIR classification was available are considered as the "events", the number of independent variables of interest (the texture parameters in this case) greatly exceeds this number of events. Naturally, by developing a model, not all of the parameters will be included. However, it is still important to ensure that the sample size (relative to the number of variables) is sufficiently large so that the assumptions made about the sample and the statistics are still valid.

Although the theoretical details of sample size restrictions will not be addressed here, they can be found elsewhere [149, 150]. One study suggests that if the number of events per variable (EPV), in the smallest group, is less than 4, then over-fitting of the data (known as Type I error) increases [151] but there are other errors can adversely affect the analysis. Under-fitting of data and paradoxical fitting (Type II and III errors respectively) will also affect the accuracy of any developed model [150, 152]. To ensure that no major errors occur in the development of the model, it is best use an EPV in the smallest group of at least 10 [150, 153, 154].

In the last chapter, much was learnt about the behaviour of the liver $R_2$ texture parameters. Some appear to correlate very well with each other and others are affected

greatly by effects other than fibrosis staging. This knowledge can be used to remove some variables from potential models in the interest of keeping an acceptable number of EPV.

### 4.2.4 Reduction of the Texture Parameter Set

Following from the analyses performed in the previous chapter, some of the texture parameters were found to be inappropriate for the quantification of fibrosis. Texture measures that were affected by confounding factors such as sensitivity to region of interest (ROI) area and sensitivity to variations within a single patient (inter-slice variations) were excluded from subsequent analyses. Some of the texture measures showed very strong correlations with other texture measures and these measures were also excluded to remove any unnecessary redundancies in the data. Redundancies were also observed in some texture measures when the pixel separations used to determine the GTSD matrices were similar. The number of pixel separations over which these similarities extend depends highly on the parameter that is being calculated (Figure 3-4). Hence, only three pixel separations for each parameter were considered – a short-range separation of 1 pixel (1.7mm), a medium-range separation of 9 pixels (15.1mm) and a long-range separation of 18 pixels (30.2mm).

After testing all parameters for the exclusion criteria, a sub-set of 15 parameters remained that were deemed appropriate to use for the subsequent model development (parameters $t_2$, $t_7$, $t_{10}$, $t_{11}$ and $t_{12}$ each calculated with pixel separations of 1, 9 and 18 pixels). Forward stepwise regression was then performed on these remaining 15 parameters so that at each step, the parameter contributing most to the predictive power of the model was included. This analysis and subsequent modelling was performed using the JMP statistical software package (SAS Institute).

## *4.3 Results*

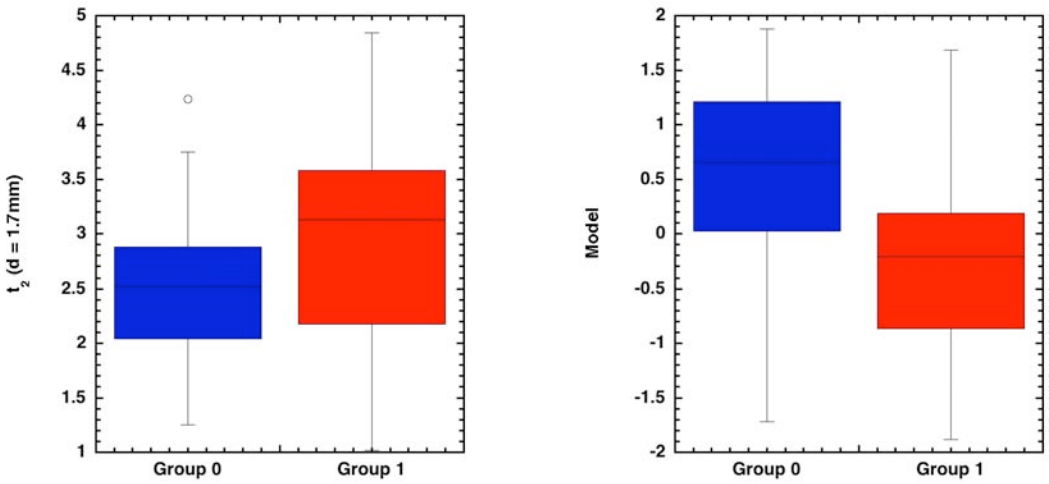
### 4.3.1 Model for Fibrosis Prediction

The forward stepwise regression generated a 4-parameter model, including parameters $t_2$ (1.7), $t_{12}$ (1.7), $t_{10}$ (15.1), $t_{12}$ (30.2) as well as an intercept term, $\beta_0$. The values calculated for the co-efficients ($\beta_i$) of this model are shown in Table 4-1. Testing the statistical significance of the model as a whole (using the whole model test) gave $p = 0.0007$.

| | Term | Estimate |
|---|---|---|
| $\beta_0$ | Intercept | 30.10 |
| $\beta_1$ | $t_2$ (1.7) | -3.13 |
| $\beta_2$ | $t_{12}$ (1.7) | 58.68 |
| $\beta_3$ | $t_{12}$ (30.2) | -57.00 |
| $\beta_4$ | $t_{10}$ (15.1) | 0.069 |

Table 4-1 – The four-parameter model (plus intercept) derived from logistic regression.

The model gives a much better distinction between the two groups of fibrosis staging for the test cohort than any single parameter. Using $t_2$ (1.7) as an indicator for fibrosis gives only a mildly significant distinction while the model derived from logistic regression gives a distinction with high statistical significance ($p = 0.0214$ and $p < 0.0001$ respectively using Wilcoxon 2-sample test). While there is a substantial overlap of the inter-quartile ranges in the case of texture parameter $t_2$ (1.7), (Figure 4-1a), the overlap is greatly reduced when considering the regression model (Figure 4-1b).




(a) (b)

Figure 4-1 – Box-plots depicting the distinction between high-grade and low-grade fibrosis.

(a) texture parameter $t_2$ (d = 1.7mm) (b) the model derived from logistic regression. The boxes indicate the inter-quartile ranges and the error bars depict the 95% confidence intervals.

The receiver-operating characteristic (ROC) curves for both $t_2$ (1.7) and $t_{12}$ (30.2) are depicted in Figure 4-2 as well as the ROC curve for the developed model. The areas under the ROC curves (± SE) were 0.64 ± 0.06 and 0.63 ± 0.56 respectively. Using the model found from logistic regression increases the area under the ROC curve to 0.75 ± 0.05.

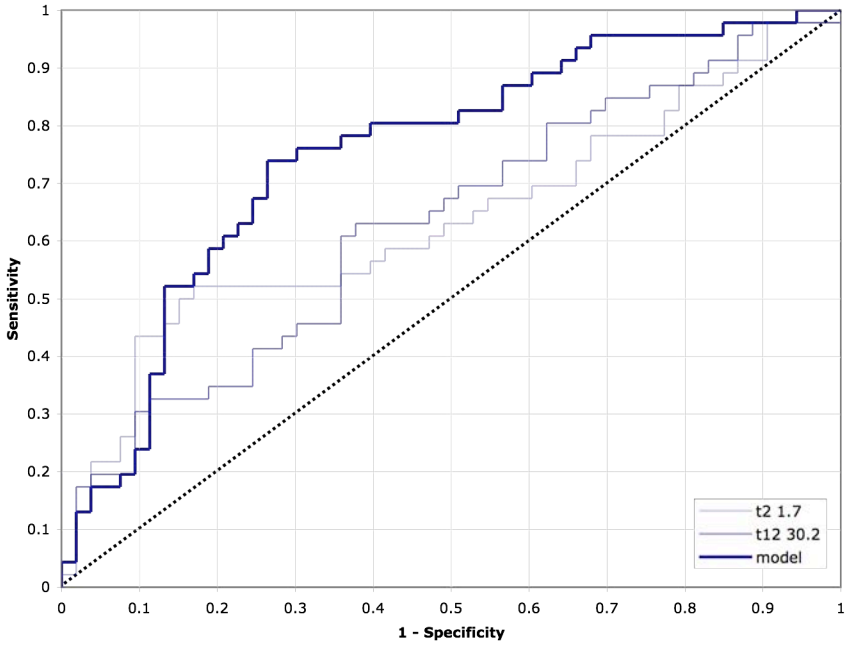



Figure 4-2 – Receiver-operating characteristic (ROC) curves for the discrimination between high and low fibrosis grades using texture measure $t_2$ (1.7), texture measure $t_{12}$ (30.2) and the whole model.

The dashed diagonal line indicates the line that has an area under the ROC curve of 0.5.

The value of the model for each subject is used to calculate the probability of that patient being in the no fibrosis group by the expression

$$P(Y=0)=\frac{1}{1-e^{g(\tilde{x})}}$$

where $g(\tilde{x})=\beta_0+\beta_1 x_1+...+\beta_n x_n,$ is the linear combination of all parameters $(x_1,x_2,...,x_n)$ derived from the model and *n* is the number of parameters in the model. So a value of zero calculated from the model is the cut-off where the model cannot distinguish between the high and low fibrosis state. A value greater than zero assigns a particular patient to the non-fibrotic group and a value less than zero indicates the presence of fibrosis. A more general expression allows the cut-off to be specified at any value of $g(\tilde{x})$ allowing the model to be assessed at another part of the ROC curve.

$$P(Y=0)=\frac{1}{1-e^{[g(\tilde{x})-g(c)]}}$$

where $g(c)$ is the value assigned as the cut-off for the model.

With the cut-off at zero, the subjects were categorised by the model as in Table 4-1 and compared to the actual fibrosis staging measured from the biopsy specimen. The sensitivity was found to be 63% and the specificity was 77%.

| | Measured Group 0 | Measured Group 1 | Total |
|---|---|---|---|
| Predicted Group 0 | 41 | 17 | 58 |
| Predicted Group 1 | 12 | 29 | 41 |
| Total | 53 | 46 | 99 |

Table 4-2 – Contingency table for the model on the test cohort using the model derived from logistic regression with a value of 0 designated as the point of distinction between high and low fibrosis.

Using the general expression with a non-zero cut-off allows the specificity and sensitivity to be calculated at a different point on the ROC curve. As a suggested example, if the cut-off is set to be 0.6 instead of 0 the contingency table will appear as in Table 4-3.

| | Measured Group 0 | Measured Group 1 | Total |
|---|---|---|---|
| Predicted Group 0 | 31 | 9 | 40 |
| Predicted Group 1 | 22 | 37 | 59 |
| Total | 53 | 46 | 99 |

Table 4-3 – Contingency table for the model on the test cohort using the model derived from logistic regression with a value of 0.6 designated as the point of distinction between high and low fibrosis.

In this case, the specificity has now increased to 80% but the sensitivity has been reduced to 58%.

### 4.3.2 Reproducibility of the Model

While the patient populations in this work do not allow for this test to be verified in a validation population, the reproducibility cohort in this study can be used to assess the reproducibility of this model when the analysis is carried out on the same patient on

different scanners on consecutive days. The $R_2$ values and the corresponding iron concentrations for each of the ten patients, for both scans, are shown in Table 4-4. By monitoring how the model and hence the fibrosis stage prediction changes over the two scans, the degree of reproducibility can be determined.

| | | Scan 1 | | Scan 2 | |
|---|---|---|---|---|---|
| Subject | Condition | $R_2$ (/s) (±SD) | LIC (mg/g dw) | $R_2$ (/s) (±SD) | LIC (mg/g dw) |
| 1 | β-Thalassaemia | 61.1 ± 17.4 | 3.00 | 65.4 ± 14.2 | 3.36 |
| 2 | β-Thalassaemia | 113.5 ± 31.5 | 8.33 | 111.5 ± 28.0 | 8.09 |
| 3 | β-Thalassaemia | 230.7 ± 69.4 | 28.28 | 196.5 ± 67.6 | 21.13 |
| 4 | Hereditary Haemochromatosis | 38.6 ± 16.4 | 1.37 | 37.6 ± 12.3 | 1.30 |
| 5 | Hereditary Haemochromatosis | 34.5 ± 10.4 | 1.12 | 31.8 ± 13.6 | 0.96 |
| 6 | β-Thalassaemia | 88.2 ± 17.0 | 5.50 | 91.4 ± 20.1 | 5.83 |
| 7 | β-Thalassaemia Intermedia | 29.1 ± 9.9 | 0.81 | 28.7 ± 12.4 | 0.79 |
| 8 | Normal | 35.5 ± 7.7 | 1.18 | 28.2 ± 10.5 | 0.77 |
| 9 | Normal | 26.2 ± 11.9 | 0.67 | 32.1 ± 13.2 | 0.98 |
| 10 | β-Thalassaemia | 49.3 ± 10.2 | 2.09 | 62.1 ± 14.9 | 3.08 |

Table 4-4 – The $R_2$ and LIC as measured from MRI in both scans for the subjects of the reproducibility cohort.

As was shown in the previous chapter, variations in slice area could impact on the accuracy of the model. With the aim of minimising these variations, the values for the model were calculated by taking the average value for the model determined in the four largest slices of each scan (Table 4-5).

| | Scan 1 | | | | Scan 2 | | | |
|---|---|---|---|---|---|---|---|---|
| Subj. | Model | P(0) | P(1) | Pred. | Model | P(0) | P(1) | Pred. |
| 1 | -0.060 | 0.341 | 0.659 | 1 | 0.141 | 0.387 | 0.613 | 1 |
| 2 | -0.531 | 0.244 | 0.756 | 1 | -0.759 | 0.204 | 0.796 | 1 |
| 3 | -0.938 | 0.177 | 0.823 | 1 | -2.435 | 0.046 | 0.954 | 1 |
| 4 | 1.449 | 0.700 | 0.300 | 0 | 1.040 | 0.608 | 0.392 | 0 |
| 5 | 1.305 | 0.669 | 0.331 | 0 | 0.618 | 0.504 | 0.496 | 0 |
| 6 | 0.914 | 0.578 | 0.422 | 0 | 1.071 | 0.616 | 0.384 | 0 |
| 7 | 1.498 | 0.710 | 0.290 | 0 | 0.927 | 0.581 | 0.419 | 0 |
| 8 | 1.142 | 0.632 | 0.368 | 0 | 0.889 | 0.572 | 0.428 | 0 |
| 9 | 0.989 | 0.596 | 0.404 | 0 | 1.191 | 0.644 | 0.356 | 0 |
| 10 | 0.733 | 0.533 | 0.467 | 0 | 0.753 | 0.538 | 0.462 | 0 |

Table 4-5 – Results from the model for both scans of each patient in the reproducibility cohort.

The model value is the average model value of the four largest slices of each scan. A model value of 0.6 was used as the cut-off for the allocation of high or low fibrosis.

Using the cut-off of 0.6 suggested, the predicted fibrosis group assignment is the same for both scans in all ten subjects in the reproducibility cohort (Figure 4-3). The majority of subjects (7/10) gave a negative test for the presence of fibrosis with three testing positive using the developed model.

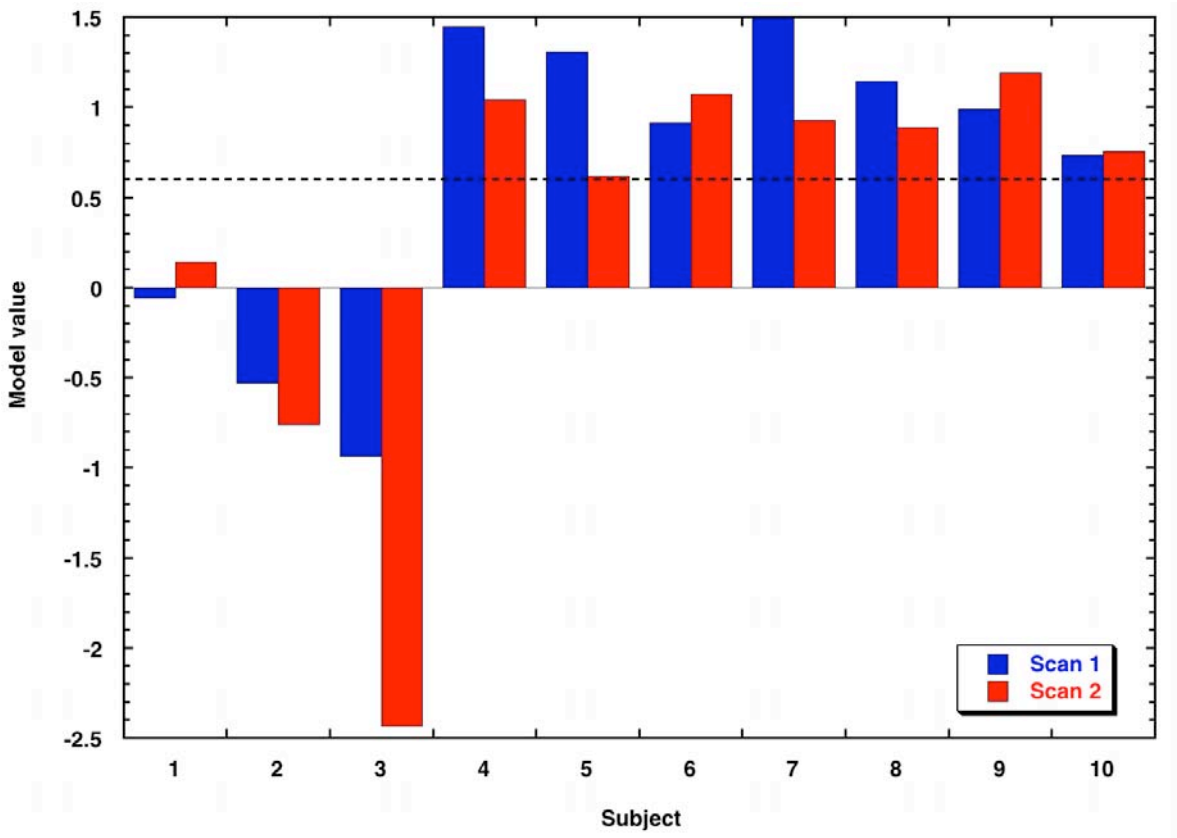



Figure 4-3 – The model value as measured for each subject in both scans.

The dashed line indicates the 0.6 cut-off value used as an example in Table 4-3.

## *4.4 Discussion*

### 4.4.1 Model for Entire Patient Set

In the process of developing the four-parameter model, other models using various numbers of parameters were investigated as candidate models for fibrosis prediction. Models with a large number of variables, while appearing statistically appropriate for this application, had associated problems. Of the parameters used in this study there are many redundancies owing to relationships between different texture measures. The introduction of more variables into the model made it more difficult to ensure that these redundancies are avoided. Also, a large number of variables included in the model violated the requirements for an appropriate EPV, potentially introducing errors in the model development. In the test cohort, the smallest group, the non-fibrotic group, has

46 events. So in this model of 4 parameters, this equates to an EPV of 11.5, which is greater than the 10 EPV required to ensure that errors of all types do not adversely affect the model [150, 153, 154]. Conversely, models with a small number of parameters, while satisfying the EPV conditions did not statistically perform as well as the models with a larger number of parameters included. The four-parameter model was determined as the most appropriate model, giving the ideal compromise between the EPV requirements and the statistical robustness.

The cut-off value for this model can be chosen to customise the analysis for the optimal specificity and sensitivity. In the assessment of liver fibrosis, the goal is to minimise the number of biopsies required to assess fibrosis. The ideal test would need to ensure that every patient testing negative for fibrosis is a true negative then only those patients that test positive will require a biopsy. A test is inappropriate if patients with fibrosis test negative and are not followed up with biopsy. To ensure the test is appropriate, a test that maximises sensitivity is required but still with sufficient specificity to ensure that biopsy can be avoided in some cases. A cut-off value for this model of 0.6 is suggested to best accommodate these criteria.

### 4.4.2 Model for Only Iron Loaded Subjects

It was demonstrated in the previous chapter that the iron in the liver appears to act as a contrast agent, enhancing the effects of fibrosis on the $R_2$ map texture. The inclusion of different patient types may adversely affect the development of the model. This model may be weighted by the effects seen in the iron-loaded subjects and hence will give unreliable results in non iron-loaded subjects. Conversely, the effects of the iron-loaded subjects in the data may be diluted in the whole patient set by the presence of non-iron loaded subjects, giving a much weaker relationship than may be present in the iron-loaded subjects.

Owing to the effects of increased sensitivity to fibrosis in the presence of iron, one would expect that a better model would arise from using only the iron loaded subjects. However, due to the very small sample size as compared to the potential number of variables, the assumptions made in the statistical methods will be invalid [150]. Future investigations should include a study on a large cohort of iron-loaded subjects to develop a model suited to iron loading.

Also seen in the previous chapter was that the type of iron loading due to a particular disease state can affect the distribution of iron in the liver. As a result, the sensitivity of a measure for fibrosis would vary depending on the type of patient in which it was measured. So there may be reason to develop a different model for each patient group. While it would be expected that a model for only the iron-loaded subjects would provide sufficient sensitivity to fibrosis, a model designed for only the non-iron loaded subjects may not be sensitive enough. If the presence of iron is required for these methods to work, perhaps texture quantification based on $R_2$ maps will not be suitable for non-iron loaded patients.

### 4.4.3 Reproducibility

It is expected that the measured fibrosis staging will not change within the 24 hours between the two scans. If these methods are indeed indicating fibrosis staging then, when performed on the same patient, the same value from the model will also not change between scans, assuming that the texture is not changing as a result of other factors. One would expect that patient motion and misregistration between the two scans would be a major factor in affecting the texture parameters measured in each scan. Recall that no biopsy specimens were obtained for any of these subjects on either scan, so there is no way of determining the actual fibrosis staging of the subject of interest.

As there are no biopsy measurements to compare with the MRI data, the results from the reproducibility cohort cannot be interpreted as a validation of the texture quantification methods for measuring fibrosis. However, there are a few observations of interest from this cohort. Firstly, for this particular model, the same fibrosis group prediction is made in both the first and second scans for every patient in the cohort. Whether indicative of fibrosis or not, these measurements are the same regardless of the scanner used or day of the scan. Secondly, the results for the healthy controls (subjects 8 and 9) show that according to this model, they are free of fibrosis and are assigned to the non-fibrotic group, as would be expected. Thirdly, there are three subjects (1, 2 and 3) that the model assigns to the fibrotic group. Two of these subjects (2 and 3) have a liver iron concentration greater than 7mg/g, the concentration at which the risk of liver damage is increased [18]. So it is reasonable to assume that of the subjects in this cohort, these two would be the most likely to have fibrosis present. The third subject that tested positive for fibrosis (1) has a liver iron concentration that is not associated with increased risk of liver damage [18]. However, this subject has β-thalassaemia

major and is on chelation therapy and could have had a liver iron concentration greater than 7mg/g in the past. Caution must be taken when interpreting these findings, as there are no biopsy measurements for these subjects, so the allocation of these three subjects to the fibrosis group is unfounded.

### 4.4.4 Comparison With Other Methods of Fibrosis Assessment

The model developed here although promising, does not perform as well as other non-invasive tests for fibrosis. Tests on biochemical markers have reported an area under the ROC curve of up to 0.9 in validation tests [49] and sensitivities and specificities exceeding those observed in this study [42]. The use of Doppler sonography and transient elastography have also demonstrated reliable tests for fibrosis [58, 67]. However, the results presented here concentrated on the discrimination of stage F0/F1 from stages F2-F4. The best results seen in other studies have been in the discrimination of F4 from lower levels of fibrosis [58]. It is expected that the discriminating power of texture measures on liver $R_2$ maps will also increase in the assessment of F4 cirrhosis. Additionally, the majority of other biochemical and radiographic tests have only looked at a specific type of patient such as hepatitis C [42] or alcoholic liver disease [43]. The model presented in this work is derived from many patients with differing pathologies. With a larger number of subjects with specific pathologies tested individually, a much more robust model is likely to be developed. Developing a model specifically for iron-loaded patients is expected to give the best results with this method.

Many diagnostic tests use parameters regarding the demographic of the patient to further improve the sensitivity and robustness. In the texture models investigated in this chapter, considerations of additional parameters such as age and iron loading mechanism in the models may have an advantageous effect on the relationships with fibrosis. However, a larger patient cohort would be required to ensure that such tests provide results that are statistically significant.

## *4.5 Summary*

The development of an appropriate method for monitoring fibrosis non-invasively could reduce the need for measurements to be made on liver biopsy specimens and minimise the sampling error attributable to the size of biopsy samples. Additionally, quantitative measures of liver fibrosis could remove the subjective nature of fibrosis grading. Texture quantification of $R_2$ maps shows promise for being indicative of fibrosis staging

but more work is required to validate them. Combining various parameters does improve the performance of a model in measuring fibrosis. However, the application of these methods for the measurement of fibrosis in the general population is likely to be limited owing mainly to the low sensitivity of $R_2$ image texture parameters to fibrosis in non iron-loaded subjects.

# Chapter 5

# The Specificity of Liver $R_2$ in the Measurement of Low Liver Iron Concentrations

## *5.1 Introduction*

Accurate non-invasive quantification of fibrosis staging is of clinical importance, as the reliance on liver biopsy would be reduced. A non-invasive quantification method would allow serial measurements and better on-going treatment options. The findings from Chapter 4 indicate that the quantification of the spatial distribution of $R_2$ in liver tissue could be useful in the non-invasive monitoring of liver fibrosis and cirrhosis, owing to the changes in iron distribution in the presence of fibrous tissue [134]. However, iron concentrations in the liver can be very low in patients with hepatitis, end stage liver disease and other severe pathologies. In the range of iron concentrations seen in these patients, the specificity of proton transverse relaxation rates to changes in iron concentration may be reduced and become influenced by other factors such as tissue hydration [155]. So, although the texture of liver $R_2$ maps is sufficiently sensitive to fibrosis in iron-loaded patients, these methods may not be appropriate in subjects with low or normal liver iron concentrations. If the measurement of liver fibrosis by using iron distribution as a surrogate is to be used in the general population, it is important to ensure that the iron is measurable with sufficient sensitivity and specificity at all liver iron concentrations.

This chapter addresses the effect that iron concentration and tissue hydration have on $R_2$ in liver tissue with end-stage liver disease and low iron concentrations. Analyses on excised liver tissue will be used to investigate the relationship between MR properties and physical characteristics of the tissue. Making measurements on *ex vivo* tissue gives information about the physical properties over a large area of liver tissue, eliminating the sampling error that exists in techniques such as liver biopsy [20, 21, 156]. Currently accepted techniques have been validated for liver iron concentrations only above 0.3mg/g dry weight [97]. The results from this study give insight into the lower limits of this relaxometry technique for liver iron quantification as well as the other factors that can complicate $R_2$ measurements at these limits.

## 5.2 Materials and Methods

### 5.2.1 Subjects

To assess how the liver $R_2$ values in a single axial MR slice are related to the physical characteristics of the tissue, a cohort of 19 subjects (15 male, 4 female, mean age 51, age range 23 – 65) all awaiting liver transplant were recruited. Of these patients, 16 were scanned prior to transplantation using the methods described in Chapter 2. Details of the subjects in the cohort are in Table 5-1. The images were processed to generate an $R_2$ map of the liver for the slice of largest cross section, using the methods described earlier. Although the time interval between the *in vivo* scan and the transplant was not constant between the patients, all patients were transplanted within a year of the *in vivo* MRI scan. The average time interval was 107 days (range 8 – 321 days).

| Initials | Sex | Age | Disease | Days |
|---|---|---|---|---|
| GR | F | 51 | Primary Biliary Cirrhosis | - |
| AS | M | 50 | Alcoholic Liver Disease | - |
| PO | M | 46 | Alcoholic Liver Disease | 108 |
| AN | M | 47 | α1 anti-trypsin deficiency | 58 |
| PA | M | 45 | Hepatitis C | 68 |
| LD | F | 65 | Hepatitis C | 223 |
| JL | M | 60 | Alcoholic Liver Disease | 123 |
| CS | M | 59 | Hepatocellular carcinoma (HCC) | 21 |
| LQ | M | 47 | Hepatitis C | 159 |
| PT | M | 38 | Auto-immune hepatitis | 73 |
| PT2 | M | 38 | Necrosis | - |
| BS | M | 54 | Hepatitis C | 244 |
| JM | M | 59 | Hepatitis B | 108 |
| RG | F | 58 | α1 anti-trypsin deficiency | 97 |
| LR | M | 63 | Blocked bile ducts | 8 |
| RM | M | 47 | Hepatitis C and HCC | 321 |
| BW | F | 50 | Hepatitis C and HCC | - |
| PB | M | 47 | Hepatitis C and HCC | 19 |
| WL | M | 57 | Hepatitis C and HCC | 42 |

Table 5-1 – Summary of the details for all patients recruited for analysis of excised tissue.

The 'Days' column is the number of days between the *in vivo* scan and the excision of the liver tissue. In the cases where no number of days is given, an *in vivo* scan was not performed.

### 5.2.2 Cut-up Procedure

Upon removal of the liver from the patient, a 2cm thick axial slice was cut from the largest cross-section of the liver. This slice was intended to be a similar cross section to that seen in the *in vivo* scans. A typical slice of liver is shown in Figure 5-1. Once

scanned, the liver cross-section was cut into 2cm x 2cm x 2cm blocks as indicated by the grid lines in the figure. The average number of blocks for a typical liver was 46 (range 28 - 68). From each block, a slice approximately 3mm thick was taken from the bottom of each block to be used for histological analysis. The remainder of each of the blocks was used for chemical iron analysis.


Figure 5-1 – Typical 2cm slice of liver obtained after excision.

The grid on which the tissue is placed is used to map the results from chemical analysis to an exact spatial location.

### 5.2.3 MR Image Acquisition and Analysis

The excised liver slice was oriented in the scanner to mimic the typical orientation of an axial slice of liver *in vivo*. This orientation was important to ensure that the axial slices in the excised scan are similar to the slices seen in the *in vivo* scan. A small water bag was held in place next to the liver cross-section to enable normalisation of the images for the purposes of the analysis. A series of spin-echo images were obtained from the liver cross-section on a 1.5T scanner, TE = 6, 9, 12, 15, 18ms and TR = 2500ms. Images were obtained for six 5mm axial slices with a 0mm gap between each of the slices.

Liver $R_2$ maps were generated as described in Chapter 2 for one slice in each of the livers. The slice closest to the board that showed no evidence of partial volume effects was chosen for this analysis. Choosing this slice is appropriate to accurately compare the $R_2$ maps with the findings from the histological analysis.

### 5.2.4 Chemical Iron Analysis

Prior to chemical analysis, the blocks were freeze-dried. The samples were weighed before and after the freeze-drying process to obtain the wet and dry weights of each block. Each of the blocks was digested using a 1:1 mixture of sulfuric (98% w/w) and nitric (70% w/w) acid. Initially, the samples were placed in 20mL of the acid mixture and placed on a low heat. As the digestion progressed, the temperature was increased and more acid was added as necessary until the digestion process was complete. The final stage of the digestion was to add 2mL of $H_2O_2$ to each sample to act as an oxidising agent to complete the digestion process. Once the digestion process was complete, each of the samples was analysed for iron using ion-coupled plasma atomic absorption spectrometry (ICP-AAS).

### 5.2.5 Model for the Prediction of $R_2$

From prior studies and MR theory, it is expected that $R_2$ will be affected by iron concentration but other factors may also play a role in affecting the measured $R_2$. These other factors become especially relevant in cases where the iron concentration is very low and any change in $R_2$ may no longer be dominated by iron. The contributions to $R_2$ can then be sub-divided in to two groups – the $R_2$ increase due to the presence of paramagnetic iron and the $R_2$ changes due to the physical composition of the tissue of interest. The latter could be affected by the amount of fat, lipids and other structures in the tissue. The ratio of the wet and dry weights of the tissue can be used as an indicator of this physical structure. A model is proposed where the predicted $R_2$ value depends on both the iron concentration (wet weight) and the wet-to-dry weight ratio (WDR)

$$R_{2,\mathrm{Pred}} = a[Fe] + \frac{b}{WDR} \tag{5-1}$$

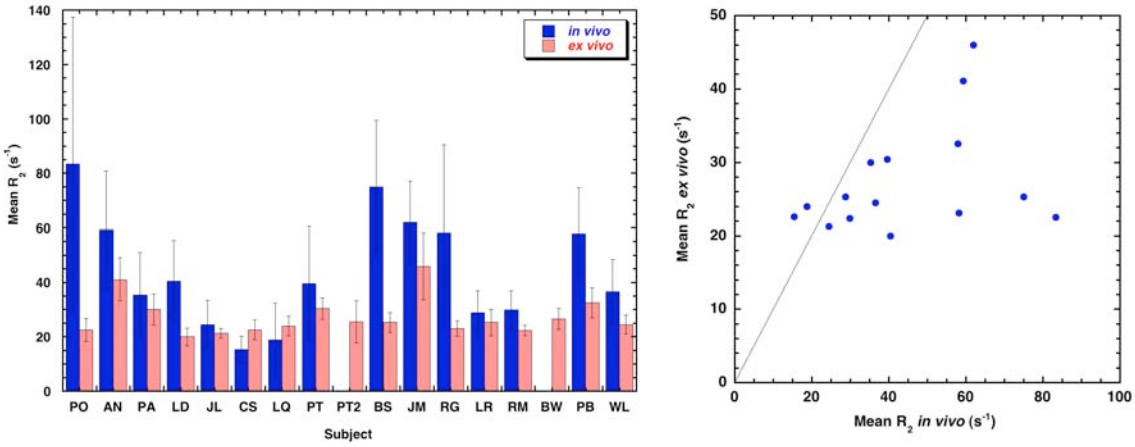
where $a$ and $b$ are constants and $\frac{1}{WDR}$ is a measure of the concentration of non-iron material in the tissue.

## *5.3 Results*

### 5.3.1 *In vivo* and *Ex vivo* Measurements of $R_2$

The measured value for $R_2$ varied between the *in vivo* and *ex vivo* scans. It is apparent that of the subjects that had an *in vivo* scan, the majority saw a decrease in $R_2$ after

excision (14/16) (Figure 5-2a). The mean change in $R_2$ (±SD) was a decrease (p = 0.004) of 16.4 (±18.7) $s^{-1}$. Spearman's rank order test showed no significant correlation between the *in vivo* and *ex vivo* $R_2$ values (p = 0.21) (Figure 5-2b).



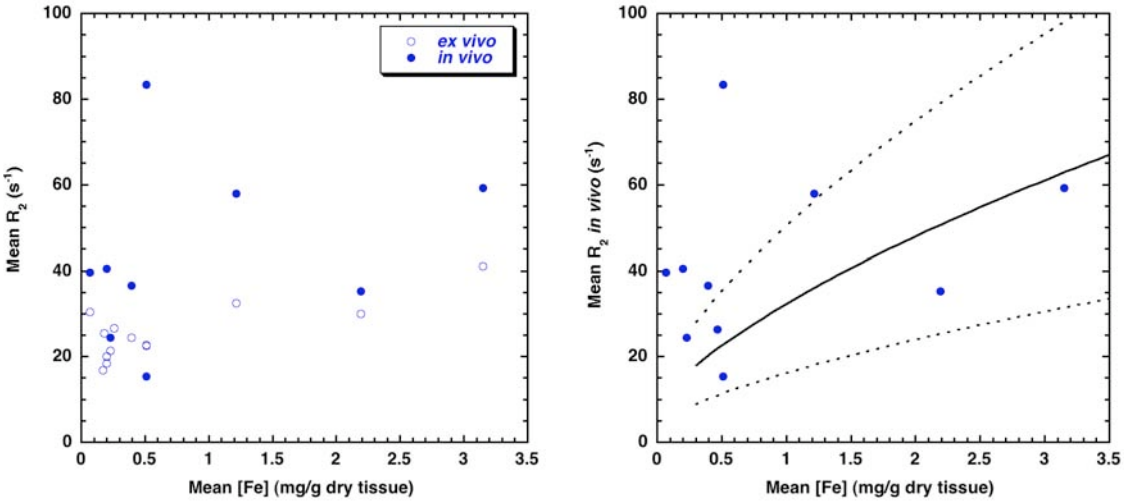

(a) (b)

Figure 5-2 – Comparison of $R_2$ measured *in vivo* and from the excised tissue.

(a) mean $R_2$ for each subject for the *in vivo* and *ex vivo* scans. The error bars indicate the standard deviation on the mean $R_2$, (b) mean $R_2$ as determined from the *ex vivo* scan plotted against the mean $R_2$ determined from the *in vivo* scan. The solid line indicates the line where *in vivo* and *ex vivo* $R_2$ values are equal.

### 5.3.2 Relationship Between $R_2$ and Iron

The measured values for $R_2$ in *ex vivo* tissue ranged from 13 to 46$s^{-1}$ while the $R_2$ calculated for the *in vivo* scans ranged from 15.4 to 83.4$s^{-1}$. The measured mean liver iron concentration was less than 0.1mg/g dry tissue in some patients and was found to be up to 3.2mg/g dry tissue in the subject with the highest iron concentration. Two patients have liver iron concentrations that lie above 1.8mg/g dry tissue – the upper limit of the normal range [157]. Both of these patients showed evidence of iron deposition in the analysis of the histology sections. Neither set of $R_2$ (*in vivo* or *ex vivo*) values showed a statistically significant relationship with dry weight iron concentration (p = 0.67 – *in vivo*, p = 0.08 – *ex vivo*) (Figure 5-3a). A large amount of scatter is evident at low iron concentrations (below 0.5mg/g dry tissue). This scatter is much larger in the *in vivo* scans. When compared to the calibration curve of St Pierre et al [97], shown in Figure 5-3b, the data for the higher iron concentrations fall within the limits of agreement. At lower iron concentrations however, the data begin to fall outside the limits of the calibration curve.




(a) (b)

Figure 5-3 – Comparison of mean $R_2$ and mean dry weight iron concentration for each excised liver.

(a) mean $R_2$ as measured in both the *in vivo* and *ex vivo* scans plotted against the mean dry weight liver iron concentration (b) mean $R_2$ as measured *in vivo* plotted against the mean dry weight liver iron concentration. Solid line is the calibration equation derived by St Pierre et al [97] and the dashed lines depict the 95% limits of agreement as stated in the same study.

The wet weight liver iron concentrations ranged from 0.02 – 0.76 mg Fe/g wet tissue. As with the dry weight iron concentrations, the $R_2$ values recorded both *in vivo* and *ex vivo* displayed no statistically significant correlation with wet weight iron concentration ($p = 0.61$ and 0.08 respectively) (Figure 5-4).

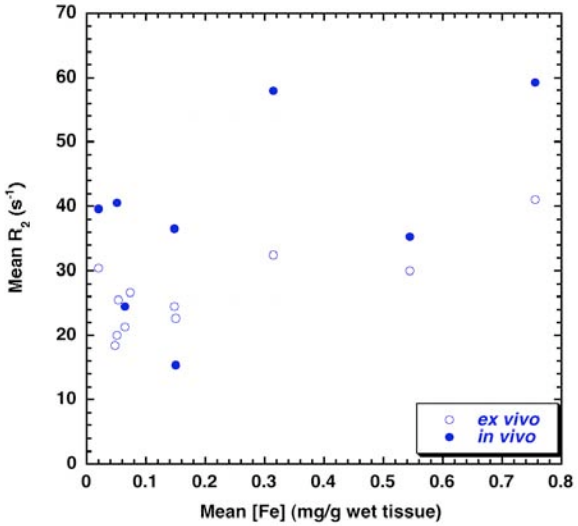
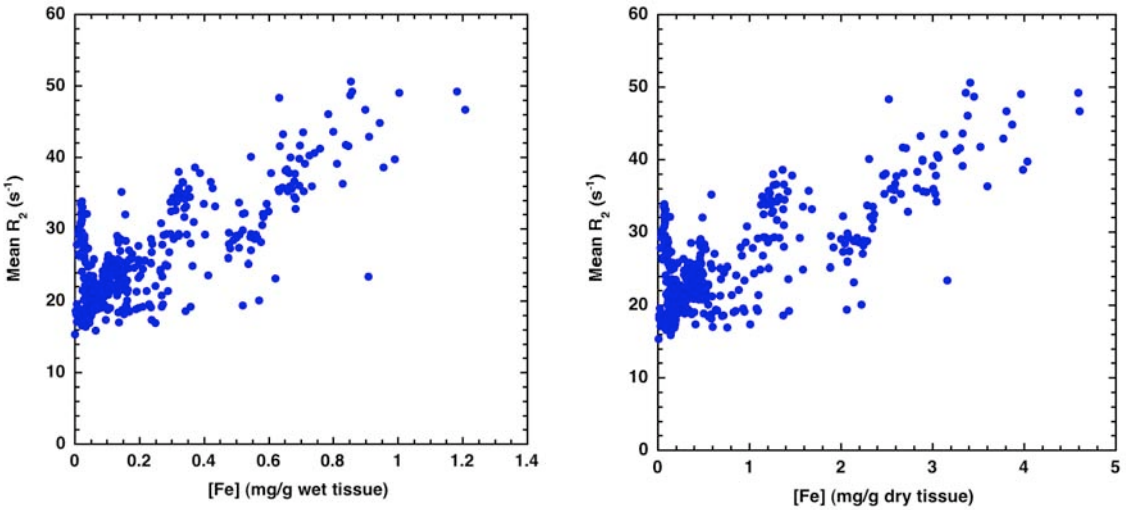



Figure 5-4 - Mean $R_2$ as measured in both the *in vivo* and *ex vivo* scans plotted against the mean wet weight liver iron concentration.

When comparing the mean liver $R_2$ and mean liver iron concentration for all the individual tissue specimen blocks, the wet weight iron concentration correlates with mean $R_2$ but there is substantial scatter in the data at low liver iron concentration ($\rho$ = 0.6025, $p < 0.0001$) (Figure 5-5a). The dry weight liver iron concentration also has scatter at low $R_2$ and the Spearman's rank order test shows a moderate and significant correlation ($\rho$ = 0.5789, $p < 0.0001$) (Figure 5-5b).




(a) (b)

Figure 5-5 – $R_2$ as measured in the individual tissue specimen blocks plotted against iron concentration.

(a) wet tissue iron concentration, (b) dry tissue iron concentration.

### 5.3.3 Relationship Between $R_2$ and Tissue Hydration

The mean wet-to-dry weight ratio (WDR) was measured for each block of tissue as an indicator of tissue hydration. For each whole liver slice, the average WDR was found to be 3.7 (range 2.7 – 4.2) and the average standard deviation in WDR within a single subject was 0.3. In both the *in vivo* and *ex vivo* determined $R_2$ values, no correlation was seen with WDR ($p$ = 0.16 and 0.77 respectively) (Figure 5-6a). With all of the blocks considered individually, the wet-to-dry weight ratio shows no significant correlation with $R_2$ ($\rho$ = 0.0085, $p$ = 0.86) (Figure 5-6b).

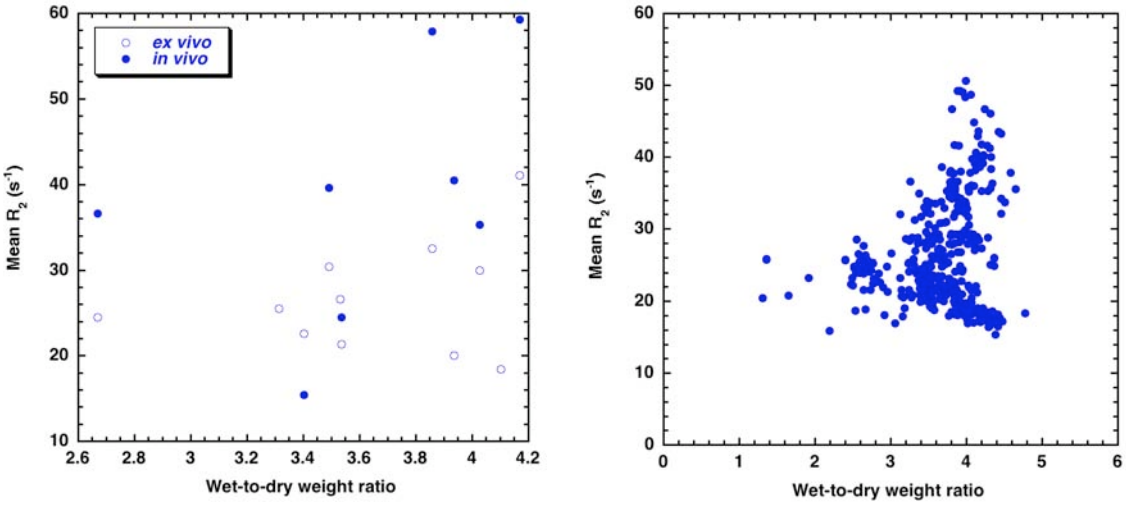



Figure 5-6 – Comparison of $R_2$ and WDR.

(a) mean $R_2$ as measured *in vivo* and *ex vivo* plotted against the mean wet-to-dry weight ratio, (b) $R_2$ as measured in the individual tissue specimen blocks plotted against the wet-to-dry weight ratio.

### 5.3.4 Sampling Variability

The iron concentration was measured for each individual tissue specimen in each liver. Using these values, the variation of iron concentration throughout a single slice of liver could be quantified. Over all of the explanted specimens, the average co-efficient of variation (CV) for a single slice of liver was found to be 44%. Similarly, the average CV in $R_2$ within a slice of liver was determined to be 15%. The contribution of iron-free tissue to the total $R_2$ of a liver was estimated from a linear fit to the data of St Pierre et al [97], for iron concentration below 3mg/g dry tissue. With this estimate, iron-free tissue was found to have a $R_2$ of $17.5s^{-1}$. Adjusting the measured $R_2$ values to only assess the $R_2$ due to the presence of iron in this tissue gave an adjusted CV of 64%.

As mean liver iron concentration increased, the standard deviation of liver iron concentration appears to increase linearly (Figure 5-7). Spearman's rank order test results ($\rho = 0.8626$, $p = 0.0001$) show a strong, statistically significant relationship between the mean and standard deviation. One data point that appears to be an outlier is included on this plot. Although this point has a very large standard deviation and a relatively small mean, there is no apparent reason to exclude this information from the analysis.

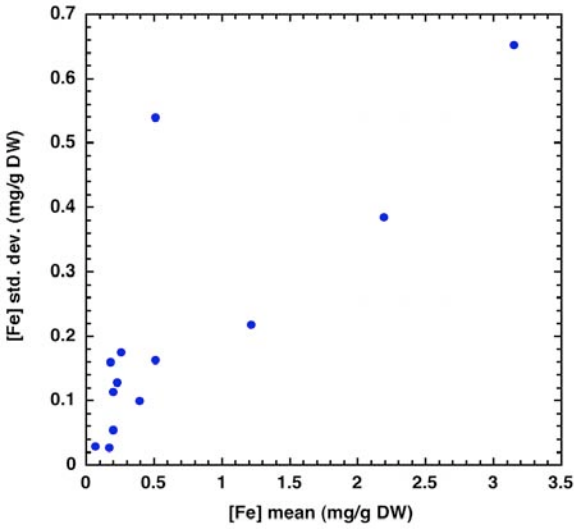
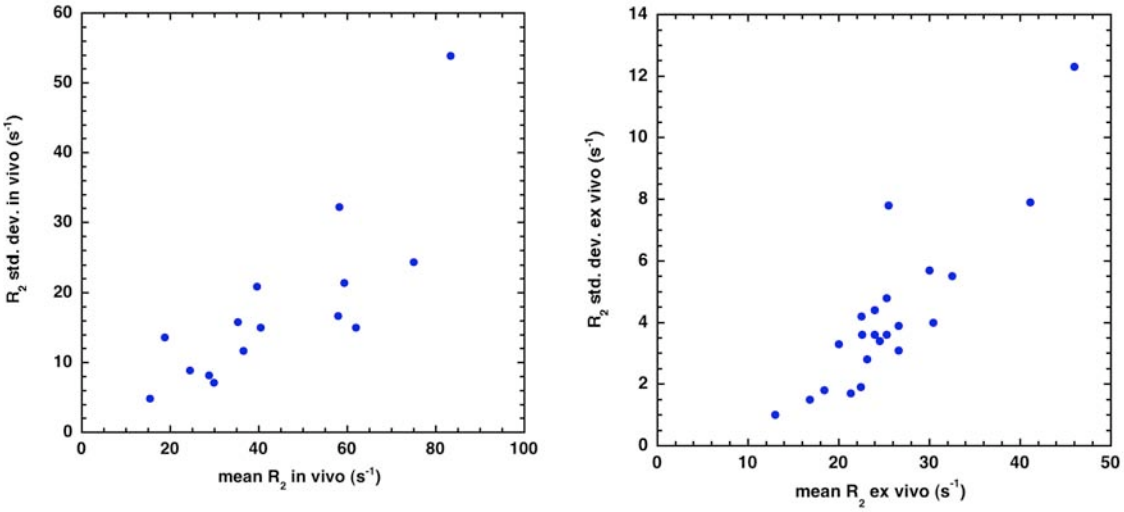



Figure 5-7 – Standard deviation of liver iron concentration plotted against the mean liver iron concentration for each excised liver.

A similar relationship was observed in the *in vivo* $R_2$ measurements of each liver, with the standard deviation of $R_2$ increasing linearly with increasing mean $R_2$ ($\rho$= 0.8365, p = 0.0001) (Figure 5-8a). A similar analysis on the WDR revealed no significant correlation between mean WDR and the standard deviation on WDR (Spearman's rank order test, p = 0.12). However, the average coefficient of variation was found to be 8%.




(a) (b)

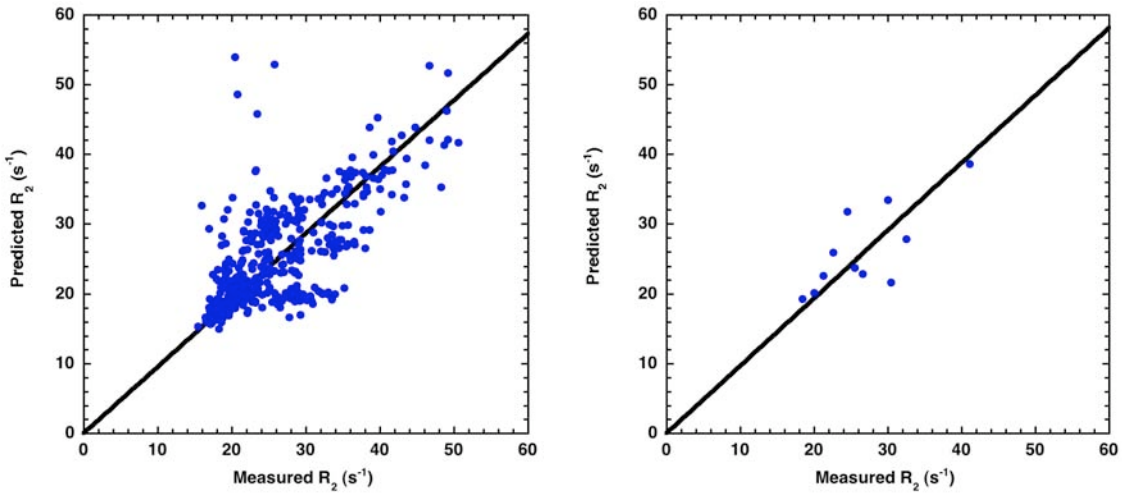
Figure 5-8 – Standard deviation of $R_2$ plotted against mean $R_2$

(a) R2 as measured from the *in vivo* scan, (b) R2 as measured from the *ex vivo* scan.

### 5.3.5 Predictive Model for $R_2$ Using Iron Concentration and Tissue Hydration

The results recorded for the wet weight iron concentration and the WDR by combining the results for each individual block over all patients were used to find a model based on Equation 5-1. Parameters *a* and *b* were found by minimising the sum of the squares of the differences between the predicted and measured *ex vivo* $R_2$. For this data set, *a* and *b* were found to be $29.1s^{-1}$(mg/g wet tissue)$^{-1}$ and $67.3s^{-1}$ respectively. A linear fit to the model had a slope of 0.95 and Spearman's rank order test showed that the predicted and measured $R_2$ are in agreement with a statistically significant correlation ($\rho = 0.67$, $p < 0.0001$) (Figure 5-9a).

The same process was applied to the average iron concentration and WDR of each subject in the cohort. A statistically significant correlation was observed between the measured and predicted *ex vivo* $R_2$ values ($\rho = 0.67$, $p = 0.02$). In this smaller set of data, the model used calculated values for *a* and *b* to be $27.8s^{-1}$(mg/g wet tissue)$^{-1}$ and $73.8s^{-1}$ respectively, in close agreement with the model found from the individual blocks (Figure 5-9b).




(a) (b)

Figure 5-9 – $R_2$ value predicted by the model plotted against the measured $R_2$.

(a) all individual block specimens (b) the average $R_2$ values for each excised liver.

Using this model, the variations in $R_2$ attributable to the iron concentration can be separated from the variations owing to the WDR. The standard deviation of the measurement of both iron concentration and WDR over all of the individual blocks can be used as an indicator of the variations in iron concentration and WDR respectively.

The variation in iron concentration accounts for a variation in $R_2$ of $30.7s^{-1}$ while the WDR accounts for variations in $R_2$ of only $3.9s^{-1}$. So the variations in $R_2$ attributable to iron are 8 times that attributable to the WDR in this cohort of patients.

## *5.4 Discussion*

### 5.4.1 Change in R2 Between *In vivo* and *Ex vivo* Scans

The lack of correlation between the average $R_2$ *in vivo* and *ex vivo* could be due to a number of factors. Firstly, owing to the movement of the liver *in vivo* it is very difficult to ensure that the $R_2$ maps prior to and post-transplant are co-registered correctly and the comparison may be between two different parts of the liver. The average $R_2$ has been shown to vary by up to 10% between different slices of a single liver [8], which may account for small discrepancies arising from misregistration issues. Although these changes in $R_2$ are inevitable, the longer times between the *in vivo* scan and the transplant will allow the liver tissue to move more and potentially enough time to change its physical structure. In this work the time between scans could not be controlled, making such variations even less predictable. While changes in physical position and shape of the liver are expected, other factors such as the liver iron concentration and structure on the cellular level may also have changed sufficiently to alter the $R_2$ value within the time between scans. Secondly, during the time that the liver is out of the body it has been shown that the $R_2$ and other relaxation rates can change significantly [158]. The changes in $T_2$ with time are most rapid within the first 120min after excision. Owing to the unpredictable nature of the excised tissue availability, the relative timing of the scan and the transplant could not be controlled and may account for part of the $R_2$ variation seen between the scans. Finally, in removing the liver from the body, the amount of heme iron present in the liver changes as blood is not necessarily retained by the liver tissue after removal. Hence, the amount of heme iron still left on the liver is not controlled between the different patients and variations in heme iron are inevitable. The signal from the heme iron manifests as part of the baseline liver signal and the loss of this iron will cause a reduction in $R_2$. In patients with low iron levels, as seen in this study, the effects of $R_2$ due to heme iron become especially important as the amount of iron in the form of heme iron becomes a more significant proportion of the total iron in the tissue.

### 5.4.2 Lower Limit of the Relationship Between Iron and $R_2$

In both the *in vivo* and *ex vivo* scans, there is a general agreement between $R_2$ and the dry weight liver iron concentration above about 0.5mg/g dry tissue. Below 0.5mg/g dry tissue, there is noticeable scatter present in the data and this is especially evident in the case of the *in vivo* $R_2$ measurements. A better correlation, although not statistically significant, is observed in the *ex vivo* case as the $R_2$ values were measured much closer to the time at which the tissue was chemically and physically analysed and no further changes to biological activity could have occurred. The shorter time interval between the two measurements allows less time for the properties of the liver tissue to change. Changes in physiology between the *in vivo* scan and transplantation are inevitable, especially in the cases where a large amount of time passed between the two scans, resulting in a much weaker correlation and more evident scatter in the *in vivo* $R_2$ measurements. The slice observed *ex vivo* is only 2cm in thickness and may not be completely representative of the slice observed *in vivo*. The existence of cirrhosis or other extreme pathology will introduce severe inhomogeneities on the macroscopic scale in both the tissue structure and iron deposition. In cases of severe inhomogeneities, co-registration becomes even more important for making accurate comparisons between the two scans and also between the *in vivo* scan and tissue analyses.

By comparison, in the calibration curve of St Pierre et al [97] in which iron concentration as determined from biopsy was compared with $R_2$ measurements *in vivo*, the biopsy and scan were performed within a few days for the majority of the patients in the study. The larger scatter in the data in Figure 5-3 is possibly a result of the longer time between the *in vivo* scan and *in vitro* analyses. The calibration was shown to be valid for iron concentrations as low as 0.3mg/g dry tissue. This limit is lower than 0.5mg/g dry tissue where the results of Figure 5-3 began to show some scatter in the $R_2$ data and no longer fell within the limits of agreement quoted for the calibration. In this study, the subjects had a number of different pathologies and were all in need of transplantation, which was not the case for the cohort used in St Pierre et al [97]. The extreme pathologies in the cohort used in this research can affect iron deposition in the tissue and other tissue structure and could be responsible for the deviation from the calibration curve.

The limits of agreement on the calibration curve are comparable to the sampling error expected from measurements of liver iron from liver biopsy specimens [97]. However, the $R_2$ values determined *in vivo* show a much larger scatter than the calibration. If the lack of agreement is indeed due to the time difference between the *in vivo* and *in vitro* analyses, then the timing of the two measurements may introduce variations in the data that dominate the error that can be attributed to the sampling error on the biopsy specimen.

The small number of subjects in this study is also a contributing factor to the apparent lack of agreement. A future investigation into similar effects with a much larger sample size will be beneficial to quantifying the agreement with existing results with a greater degree of certainty. All patients in this cohort had some form of extreme pathology for which transplantation was required. The ultra low iron concentrations most likely result from the prevalence of cirrhosis, where iron will preferentially accumulate in the nodules [134]. In these extreme pathologies, the $R_2$ is quite possibly affected more by physical changes in the tissue rather that the presence or absence of iron.

In organs such as the brain where $T_2$ is comparable to that of non iron-loaded liver [159], longer echo times have been used to obtain a better description of the signal decay. At such low $R_2$, the selection of echo times used in this study may not allow for accurate modelling of the decay curve. In future investigations into effects on $R_2$ in livers of low iron concentrations a more appropriate set of echo times would be advisable.

### 5.4.3 Effect of Tissue Hydration on the Relationship Between Iron and $R_2$

The scatter in $R_2$ observed at low iron concentrations indicates that the variation in $R_2$ must be influenced by factors other than iron. Factors such as tissue type, tissue structure and WDR can all act to cause changes in $R_2$. However, all of these effects will act simultaneously on the $R_2$ of a tissue of interest and cannot be considered independently. These other factors become very important at low iron concentrations where the changes in $R_2$ may no longer be dominated by changes in iron concentration. By considering the variation in $R_2$ due to both iron concentration and WDR as in the model of equation (5-1), very good agreement is observed between the model and the measured $R_2$ values. The analyses of Chapter 4 rely on the fact that $R_2$ gives a reliable

and specific indication of the distribution of iron within the liver, which is affected by the presence of fibrosis. The fact that the measurement of $R_2$ at low iron concentrations is significantly affected by both iron and tissue hydration may explain the lack of sensitivity of liver $R_2$ texture to fibrosis in the absence of iron-loading.

For liver $R_2$ to be dominated by iron concentration, the measured $R_2$ must exceed both the contribution of the iron-free tissue to $R_2$ ($17.5s^{-1}$) and the variation owing to tissue hydration ($3.9s^{-1}$). So for $R_2$ values above $21.4s^{-1}$, the effect of iron on the $R_2$ of liver tissue would be expected to dominate the effect of tissue hydration. According to the calibration curve of St Pierre et al [97], an $R_2$ of $21.4s^{-1}$ corresponds to a liver iron concentration of approximately 0.5mg/g dry tissue. This liver iron concentration is in agreement with the breakdown of the relationship between liver iron concentration and $R_2$ observed in this study.

The model including both iron concentration and WDR raises the correlation coefficient observed when just considering the iron concentration only, even though no correlation was observed in the WDR data. The inclusion of both variables appears to correct for the effect of tissue hydration on the $R_2$ values. As a result, the scatter observed in $R_2$ at low iron concentrations is not apparent in the model. Such a model could be used clinically where accurate mapping of $R_2$ for low iron concentrations is important and effects due to changes in tissue need to be removed. This model, although providing a good fit to the data in these samples, needs to be verified in a test set of patients. The WDR does not appear to vary significantly over a slice of liver with an average standard deviation of 0.3. To obtain a larger sample size in which to verify this model, it may be practical to repeat these analyses on patients requiring biopsy and obtaining iron and WDR measurements on a single biopsy specimen rather than explanted tissue.

### 5.4.4 Sampling Variability

The analysis of many sub-samples in a single slice of liver allows for a quantification of the typical iron concentration variation in liver tissue. The average CV of 44% observed for the specimens in this study is comparable to the 40% seen previously [20, 21]. The linear relationship observed between the mean iron concentration and standard deviation was also observed in these studies. This variation however, is not reflected in the iron contribution to the $R_2$ measurements where the average CV was 64%. At the low iron concentrations seen in this work, the changes in $R_2$ do not only result from

changes in iron concentration but by other effects such as tissue hydration. The variations in $R_2$ additional to the 44% owing to inhomogeneous iron concentrations can be attributed to the variations in WDR and other physiological variations throughout the tissue.

## 5.5 Summary

Previous studies have shown that mean liver $R_2$ is a good indicator of mean liver iron concentration in the management of iron overload disease and has been tested and verified in cases of normal up to very high iron loading. However, this relationship starts to break down at iron concentrations below 0.5mg/g dry tissue. Care needs to be taken in interpreting $R_2$ values in these cases as the absence of excess iron contributing to $R_2$ gives way for other factors to have significant influence. One such factor is the change in WDR of the tissue. While WDR on its own does not display any relationship with $R_2$ in the liver tissue, including its effect in the analysis does give a much better indication of the changes in $R_2$ seen in liver tissue with very low to normal iron concentrations. A number of difficulties arise in accurately quantifying $R_2$ at these iron concentrations due to the significant contribution of heme iron in low iron tissues and the inherently slow relaxation rates. These issues observed in the presence of low iron concentrations will reduce the specificity of $R_2$ to liver iron concentration and possibly be responsible for the decreased sensitivity of liver $R_2$ texture to fibrosis.

# Chapter 6

# Assessment of Magnetic Susceptibility Artefacts in Proton Relaxometry of Synthetic Phantoms

## 6.1 Introduction

There are many different techniques available that exploit MRI to make measurements of tissue iron concentrations [8, 92, 93, 95]. These techniques can be separated into two major areas of study, $T_2$ (or $R_2$) methods based on images obtained from spin-echo sequences and $T_2$* (or $R_2$*) methods that are based on gradient-echo imaging modalities. Both have shown very good agreement with physical measurements of iron concentration in the liver [97, 118] and $T_2$* methods have been used as an indicator of cardiac iron concentrations [95]. The 180º pulse that is used in the spin-echo sequence refocusses all dephasing except for the variations due to the microscopic effects such as those on the order of the diffusion length of water within the echo time, about 10μm [131]. In the gradient-echo sequence, the rephasing, due to a reversal of the applied field gradient, does not eliminate the relaxation mechanisms resulting from time-independent effects.

While $R_2$ methods are sensitive to only the magnetic field variations on lengths less than the diffusion length of water, $R_2$* has additional relaxation mechanisms including the effects of macroscopic field gradients. These can be generated by inhomogeneities in the iron distribution as well as from regions where dramatic changes in magnetic susceptibility occur such as at organ boundaries. These susceptibility interfaces could be useful as they enhance edges due to spatial concentration changes of iron but may be affected by neighbouring tissue, vessels or other structures.

The work in this chapter quantifies the effect and extent of the artefacts generated by sudden changes in magnetic susceptibility in a synthetic model. Results from the experiments were compared with the effects expected from theoretical models using three different geometries. Theories about the enhancement at susceptibility interfaces can then be used to quantify the effects of different geometries and sizes on $R_2$ and $R_2$* based relaxometry methods.

## 6.2 Theoretical Description of Magnetic Field Gradients in Simple Geometries

### 6.2.1 Sphere

The magnetic field of a sphere of uniform magnetic susceptibility acts as if the field is due to an ideal magnetic dipole located at the centre of the sphere. At any given location ($r$, $\phi$) from the centre of the sphere, the magnetic field experienced will be perturbed by the field of the spherical object, as given by the following expression [86].

$$\Delta B_{z,\text{Sphere}}(r,\phi) = \frac{\Delta\chi}{3+\Delta\chi}\left(\frac{R_S}{r}\right)^3(3\cos^2\phi - 1)B_0 \qquad \text{for } r \geq R_S$$

$$\Delta B_{z,\text{Sphere}}(r,\phi) \approx \frac{\Delta\chi}{3}\left(\frac{R_S}{r}\right)^3(3\cos^2\phi - 1)B_0 \qquad \text{for } \Delta\chi << 1$$

where $R_S$ in the radius of the spherical object, $B_0$ is the magnitude of the applied magnetic field and $\Delta\chi$ is the difference in magnetic susceptibility between the sphere and the surrounding medium. The $T_2$* decay of the detected signal is enhanced by the gradient of this perturbation

$$\vec{G}_{\text{Sphere}}(r,\phi) = \nabla(\Delta B_{z,\text{Sphere}}(r,\phi))$$

The field lines of an ideal magnetic dipole are shown in Figure 6-1. Qualitatively, it can be seen that the gradients due to the z-component of the magnetic field will be highest at the poles of the dipole. Additionally, there will also be a significant contribution to the signal decay around the equator of the sphere. It should also be noted that at some angles, there would be no additional signal decay caused by the field of the sphere. This angle is the double cone formed when $\phi$ is such that $3\cos^2\phi - 1 = 0$ (54.7° from the direction of $B_0$), indicated by the dashed line in Figure 6-1.

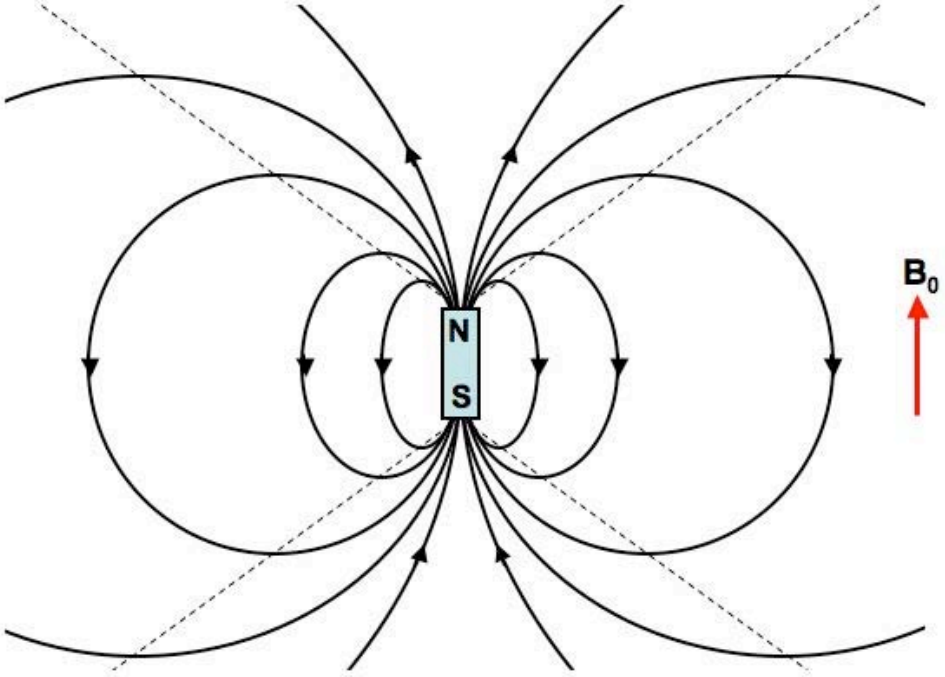
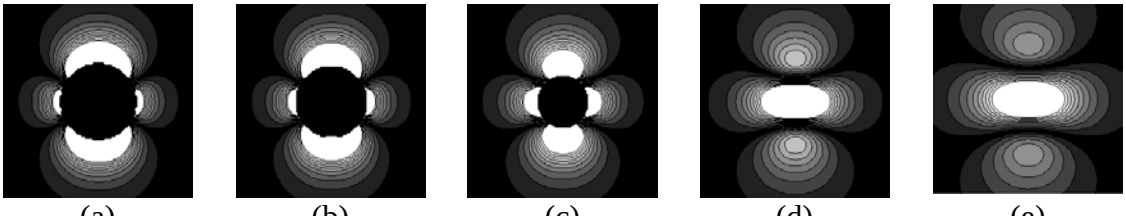



Figure 6-1 – Diagram of the magnetic field lines from a theoretical dipole.

The dashed lines indicate the double-cone where the component of the magnetic field in the direction of the applied field is zero.

The series of images in Figure 6-2 are contour plots of the field gradient of a sphere (radius 20mm) for various slices (5mm thickness) from the equatorial slice through to slices past the physical extent of the sphere. These plots clearly show the features that were expected from the qualitative consideration. Intense lobes exist at the poles of the sphere, with additional enhancement observable around the equator. The double-cone where no gradients are observable due to the z-component of the field is also evident. Notably, the distance to which these gradient effects exist appears to extend well beyond the vicinity of the sphere. This extent however, does depend highly on the angle at which a voxel of interest is situated from the z-axis of the sphere.



(a) (b) (c) (d) (e)

Figure 6-2 – Contour plots (for 5mm thick slices) of the field gradient of a sphere of 20mm radius at varying distances from the central slice

(a) central slice, (b) 7.5mm away from central slice, (c) 15mm away from central slice, (d) 22.5mm away from central slice, (e) 30mm away from central slice.

The derivation above depicts the effect of a sphere of 2cm radius. However, the observed effects will depend on the size of the sphere. Figure 6-3 shows how the magnitude of the field gradient is expected to fall off in the equatorial slice from the north pole of the sphere, moving away from its surface in the positive z-direction (as indicated by the arrow). The quantification of signal enhancement is expressed in arbitrary units as an exact calculation of $R_2$* will depend on other machine-dependent constants.

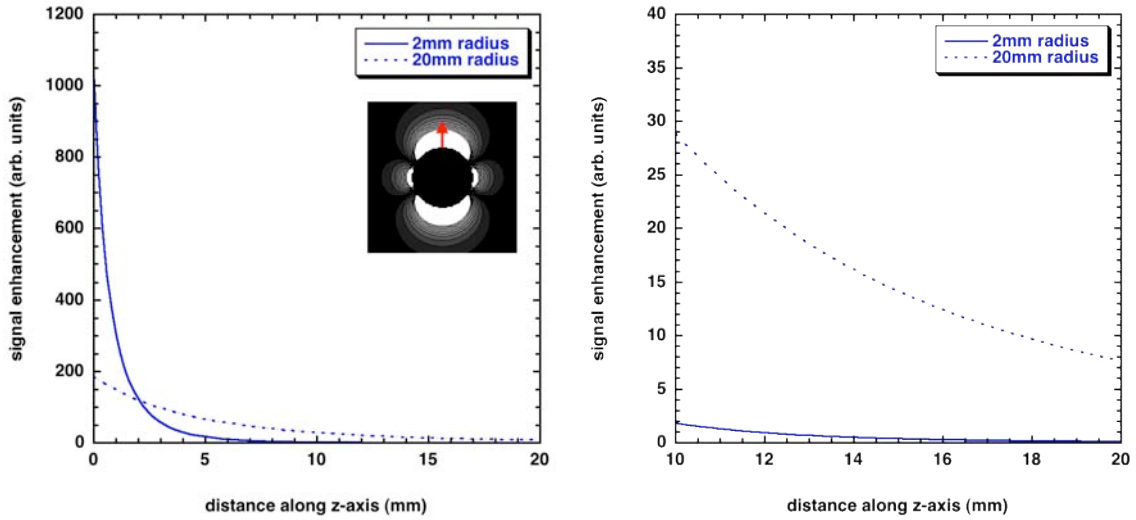



Figure 6-3 – $R_2$* signal enhancement ($R_2'$) plotted against the distance from the surface of the sphere in the positive z-direction from the "north pole" of the sphere in the central slice.

The direction of the distance axis in the plots is indicated by the arrow in the inset. Signal enhancement is indicated for spheres of radius 2mm (solid line) and 20mm (dashed line). (a) signal enhancement from the surface of the sphere to 20mm away from the surface, (b) detail of signal enhancement between 10 and 20mm away from the surface.

Figure 6-3a shows the previously considered case where the sphere radius is 20mm (dashed line) as well as the result for a sphere of 2mm radius (solid line). For the 20mm sphere, at a point 20mm away from the edge of the sphere, the magnitude of the gradient has dropped to about 10 from a value of 180. In contrast, the sphere of a 2mm radius, as shown in Figure 6-3b, the magnitude of signal enhancement at the edge of the sphere is almost 1000 - 5 times that seen for the 20mm sphere - but at a point only about 1cm away, the magnitude has effectively dropped to zero. So the size of the sphere not only affects the gradients in the vicinity of the sphere, but also the rate at which the

gradients fall off with distance from the magnetic susceptibility interface. These findings extend to clinically applicable geometries as well. For example, a highly iron loaded liver could have a magnetic susceptibility comparable to that of air [71] and an effective radius of curvature of about 200mm. At the interface of such an organ with other tissue, the magnitude of signal enhancement would be 11 but at a point 20mm away will be 7, a decrease of less than 50%.

### 6.2.2 Infinite Cylinder

Unlike the case of a sphere, the effect of a finite cylindrical object cannot be solved using analytical methods. However, the theoretical case of on infinitely long cylinder can be solved analytically giving the following expression for the perturbation of the z-component of the magnetic field [86].

$$\Delta B_{z,\mathrm{Cyl}}(r,\phi) = \frac{\Delta\chi/2}{1+\Delta\chi/2}\left(\frac{R_c}{r}\right)^2 B_0 \cos 2\phi \sin^2\alpha \qquad \text{for } r \geq R_C$$

with $\alpha$ representing the angle between the applied magnetic field and the long axis of the cylinder. For the case where the cylinder is perpendicular to the applied field, this expression reduces to

$$\Delta B_{z,\mathrm{Cyl}}(r,\phi) = \frac{\Delta\chi/2}{1+\Delta\chi/2}\left(\frac{R_c}{r}\right)^2 B_0 \cos 2\phi \qquad \text{for } \alpha = 90^{\circ}$$

$$\Delta B_{z,\mathrm{Cyl}}(r,\phi) \approx \frac{\Delta\chi}{2}\left(\frac{R_c}{r}\right)^2 B_0 \cos 2\phi \qquad \text{for } \Delta\chi << 1$$

In reality, an expression such as this would apply for modelling magnetic field effects at a point sufficiently far from the ends of a relatively long cylinder. As shown by Reeder et al [130], this approximation can be used to model the effects of $T_2$* decay in the vicinity of blood vessels.

The effects of $T_2$* decay at the ends of a cylinder cannot be described using analytical methods. For a cylinder aligned parallel to the $B_0$ field, conceptually one could imagine the field lines generated from such an object will be similar to that of a bar magnet. Qualitatively, this would cause very large gradients at the poles of the cylinder but relatively small gradients along the sides of the cylinder.

As in the spherical case, the perturbation to the $T_2$* decay is caused by the gradient field of the cylinder which is given by

$$\vec{G}_{\mathrm{Cyl}}(r,\phi) = \nabla(\Delta B_{z,\mathrm{Cyl}}(r,\phi))$$

Qualitatively, the shape of the field gradients in the cylinder (Figure 6-4) is similar to that of the sphere (Figure 6-2). However, the size and extent of the gradients in the plane orthogonal to the field direction are larger than that of the sphere. This difference is attributable to the lack of curvature through the plane of the image.

(a) (b)

Figure 6-4 – Contour plot of the field gradients for a cylinder aligned orthogonal to the field.

(a) cylinder of 10mm radius , (b) cylinder of 20mm radius.

## *6.3 Materials and Methods*

### 6.3.1 Phantom Design Considerations

The nature of this experiment has a number of distinct requirements. Firstly, there must be a magnetic susceptibility difference between the object of interest and the surrounding medium to ensure a distinct susceptibility interface at which the field gradients predicted by the theory can be tested. Secondly, the relaxation rate of the sample surrounding the object of interest must be in the range where it can be mapped using the available software. A relaxation rate that is too high will cause problems in the modelling of the exponential decay, as the signal will decay too rapidly to be correctly modelled. Using shorter echo times could give a better representation of the signal decay but with the currently available MR unit software and hardware, these echo times are impractical with a spin echo sequence. On the other hand, lower relaxation rates will require much longer echo times to accurately fit the exponential decay curve. Finally,

the medium must be homogeneous so that the effects from susceptibility interfaces are only due to the object suspended in the medium.

The final design criterion was to design a method to hold the object in a fixed location in the medium. Any motion of either the object or the medium would result in undesirable artefacts that could complicate the effects of the susceptibility interface. To overcome this issue, a 15-20% w/w gelatine was chosen as the medium as it gave sufficient rigidity to minimise the motion of the object within the phantom [160].

### 6.3.2 Phantom Construction

In this study, phantoms were constructed by suspending air-filled objects in gelatine. To ensure that the gelatine would give the desired signal decay in the MRI analyses, the gels were doped with a $MnCl_2$ solution of approximately 1mM concentration. The first phantom had a 4cm diameter ping-pong ball suspended in the centre of the container (Figure 6-5a). The second used a test tube 10cm in length and about 1.5cm diameter suspended in the gelatine in the same manner (Figure 6-5b). In both cases, the objects of interest were suspended such that they were sufficiently far from the edges of the box to ensure that any effects that may arise from the air-container interface are negligible. Additionally, the test-tube and ping-pong ball were chosen under the assumption that the material from which they were constructed met three criteria. Firstly, the material needed to be sufficiently rigid so that it held its shape in the gelatine. Secondly, the material needed to be sufficiently thin so as to minimise any effects that may arise from the material itself. Finally, the object needed to be sealed so that no gelatine was allowed to flow into the object prior to setting.

### 6.3.3 MRI Image Acquisition and Analysis

Once the gelatine in the phantoms had set, MR images were obtained using two different pulse sequences. The first was a single spin-echo sequence generating a series of spin-echo images at echo times of 6, 9, 12, 15 and 18ms using a repetition time of 2500ms. Secondly, a gradient echo sequence using the same echo times and a repetition time of 2000ms gave a series of 5 gradient-echo images. For both geometries a slice thickness of 5mm was used but the slice spacing was set to 5mm for the tube while in the phantom with the ball, a slice spacing of 2.5mm was used (Figure 6-6). Initially the tube phantom was set up so that the tube lay along the direction of the applied static field $B_0$ and coronal slices were taken. The tube was then scanned a second time with

the tube this time aligned so that it was orthogonal to the applied field. In this instance, sagittal slices were used to obtain a series of images of the circular cross sections of the cylinder. With the cylinder in this geometry two sets of scans were obtained using two different phase-encoding directions.

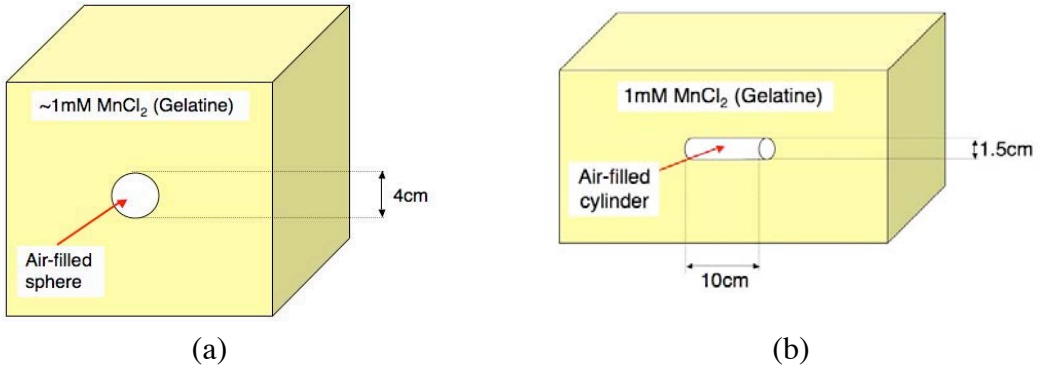



Figure 6-5 – Orientation of the air-filled object in the $MnCl_2$-doped gelatine phantoms.

(a) Air-filled sphere suspended in gelatine, (b) Air-filled cylinder suspended in gelatine.

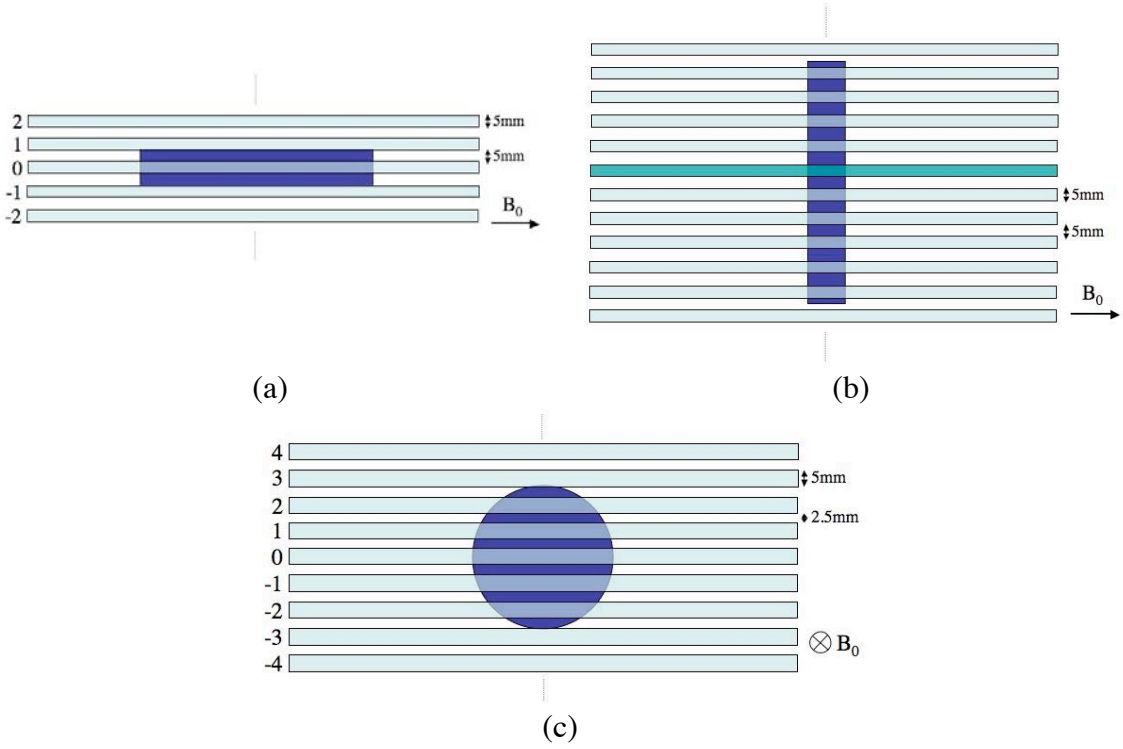



Figure 6-6 – Orientation of the slices in the MR images

(a) air-filled cylinder aligned parallel with the applied magnetic field (b) air-filled cylinder aligned perpendicular to the applied magnetic field. The shaded slice is the location of the slice used for relaxometry analysis (c) air-filled sphere. The direction of $B_0$ is indicated by the arrow.

$R_2$ and $R_2$* maps were generated from the raw images using the methods stated in Chapter 2. Spin-density projection and gain correction were not used in the analyses of these phantoms. $R_2'$ maps were calculated by pixel-wise subtraction of $R_2$ from $R_2$*.

## *6.4 Results*

### 6.4.1 Sphere

The spin-echo and gradient-echo images for the sphere phantom are depicted in Figure 6-7. The images shown are of the equatorial slice and were obtained using echo times of 6, 12 and 18ms. In the spin-echo images, the outline of the ball is quite obvious and distinct, even though the shape appears to be distorted somewhat. The gradient-echo, similar to the spin echo images shows some distortion as well but to a much larger degree than the spin-echo case. The most striking feature of the gradient-echo images is the loss of signal intensity in the later echoes beyond the boundary of the sphere. This region appears more hypo-intense as the echo time increases.

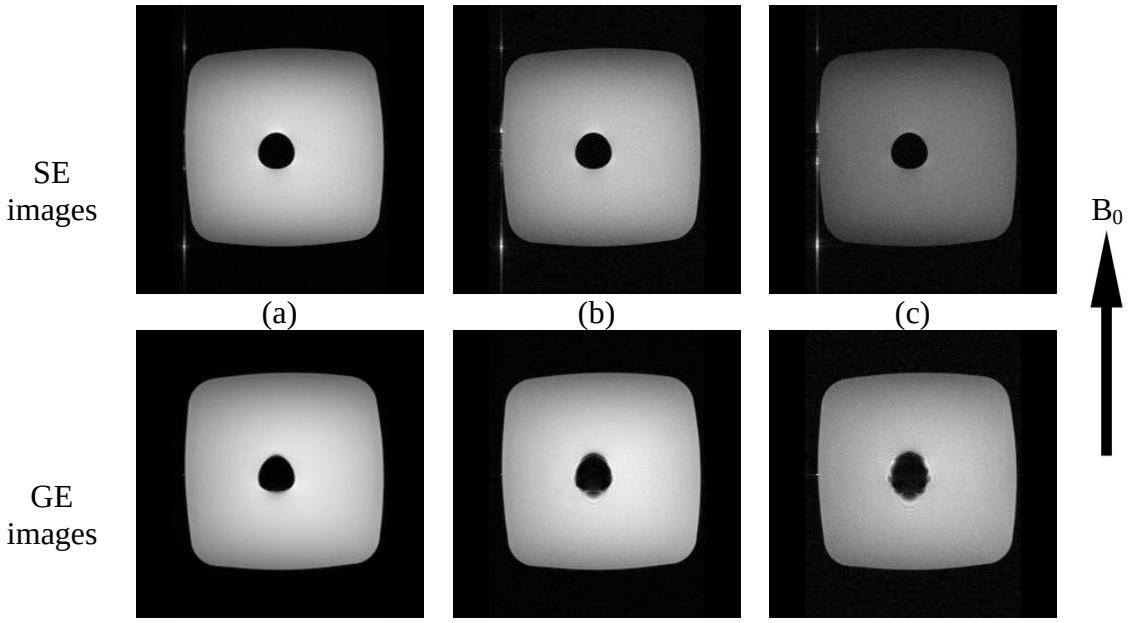



Figure 6-7 – Raw images of central slice for the air-filled sphere

(a) TE = 6ms, (b) TE = 12ms, (c) TE = 18ms. The top row of images were obtained from the spin-echo scan, the images on the bottom row were obtained from the gradient-echo scan. The static-field direction is indicated by the arrow.

After the images were processed with the relaxometry software, $R_2$ and $R_2$* maps were generated from the spin-echo and gradient-echo images respectively. Simple image subtraction was then applied to generate $R_2'$ maps of the region surrounding the ball

(Figure 6-8). The $R_2$' map for the slice that includes the equator of the ball is shown in Figure 6-8a. The subsequent figures are the $R_2'$ maps for the slices moving upward from the equatorial slice (Figure 6-8b-c). Moving further upwards, Figure 6-8d and Figure 6-8e depict the $R_2'$ maps at the slices above the top of the ball, It can be seen that as the slices move further and further from the top of the ball, the $R_2'$ value drops closer to the background $R_2'$. A similar result is found when considering the slices that lie below the equatorial slice of the ball.

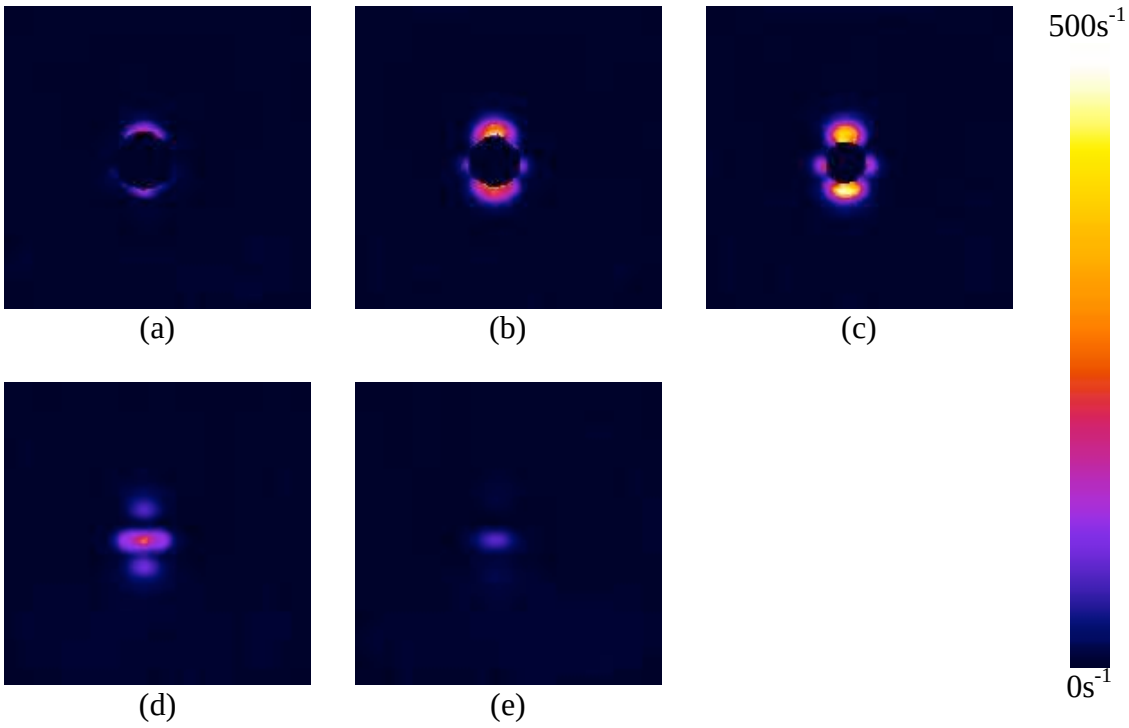



Figure 6-8 – $R_2'$ maps for the air-filled sphere

(a) the central slice, (b) 7.5mm above central slice (c) 15mm above central slice, (d) 22.5mm above central slice, (e) 30mm above central slice.

The most intense $R_2$* enhancement in each slice is observed at the surface of the sphere at what can be considered the north and south poles of the sphere. The $R_2'$ at these points initially increase for the slices adjacent to the central slice and then decrease as further slices are considered. The maximum $R_2'$ is not observed at the equator but rather the slices above the central slice. A similar result is seen both above and below the central slice (Figure 6-9).

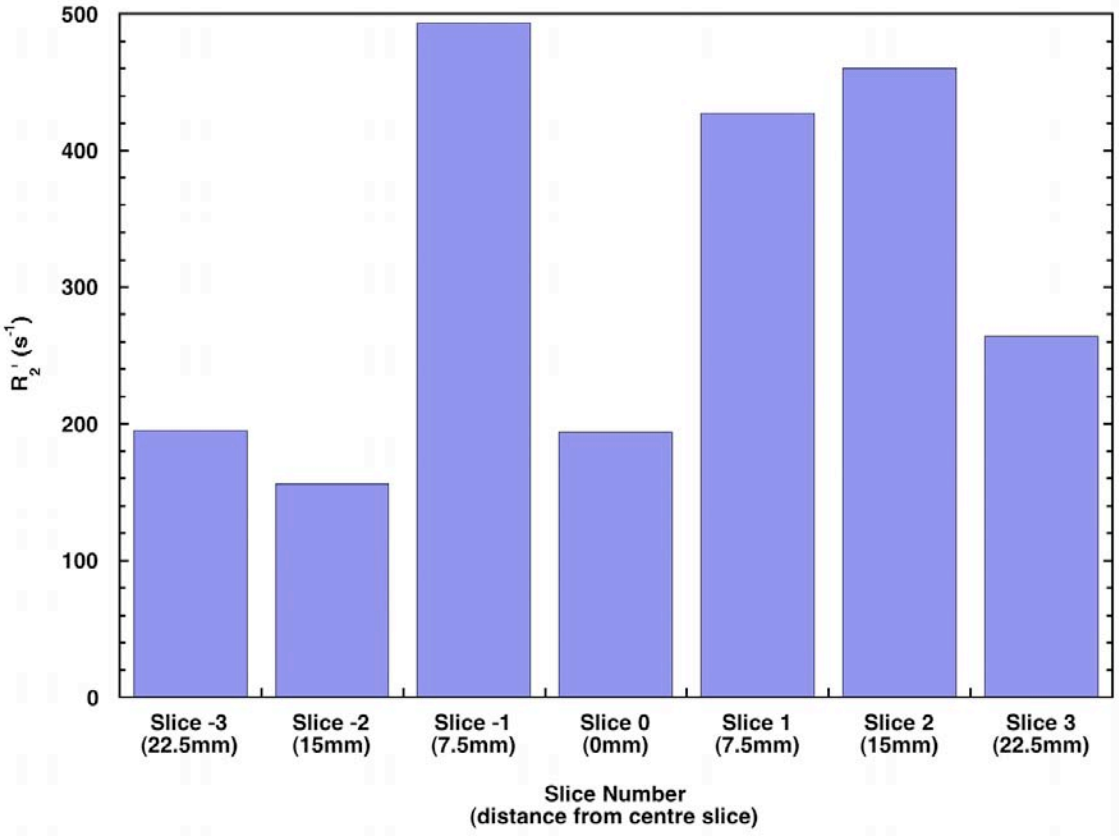



Figure 6-9 – The $R_2'$ measured at the point at the surface of the "north pole" of the sphere for the central slice as well as slices above and below the central slice.

For comparison, Figure 6-10 shows the effective $R_2'$ enhancement versus the distance of the slice from the equator as predicted by the theory for slice thicknesses of 5mm, 10mm, 20mm and the theoretical case of 0mm where the slices can be considered to be purely 2 dimensional. Note that these plots only consider signal enhancement due to the effects of the susceptibility interface and not other factors that may contribute to $R_2'$.

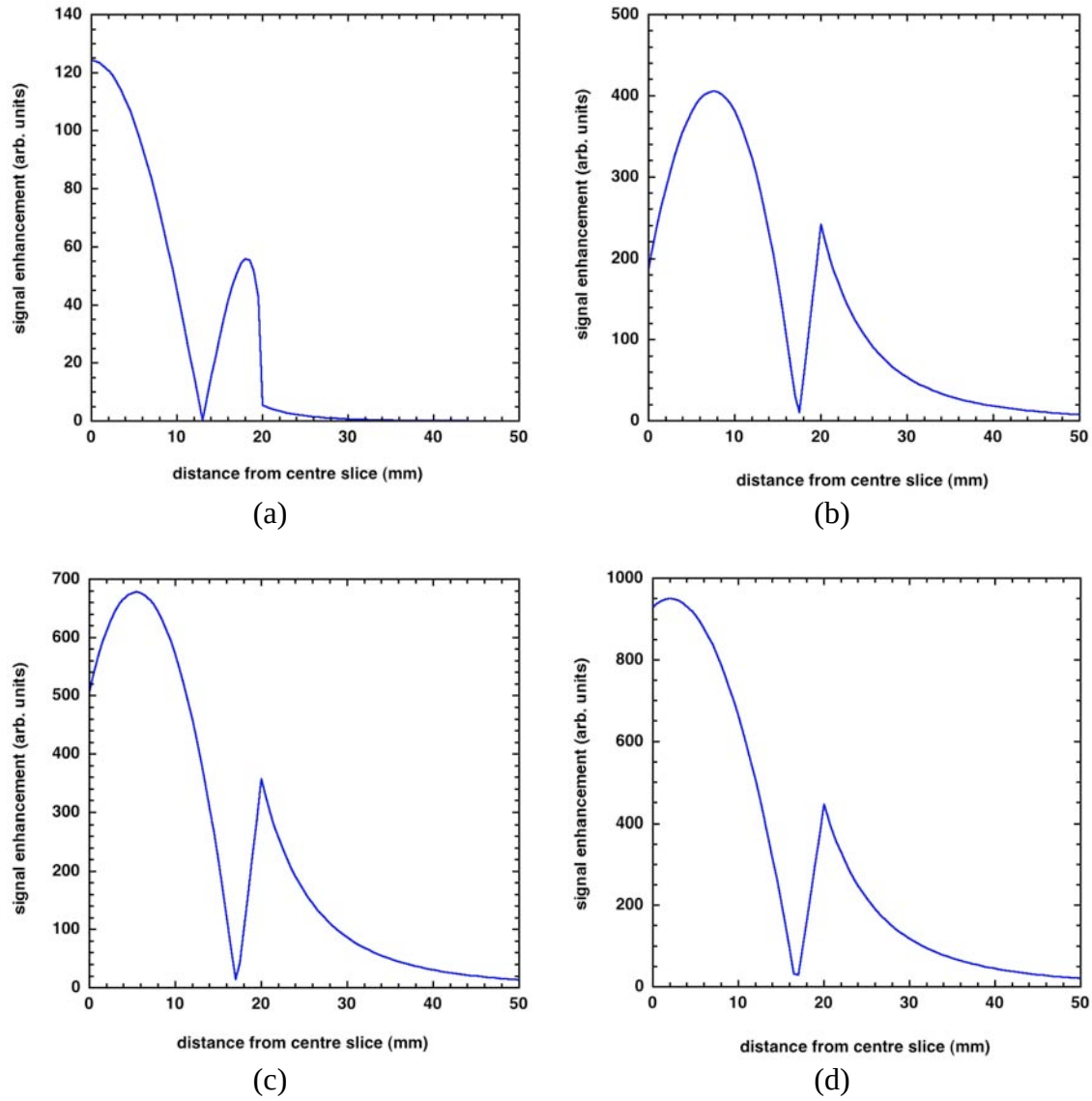



Figure 6-10 – Signal enhancement at the surface of the sphere, at the "north pole", as predicted by theory plotted against the distance from the centre slice.

Slice thicknesses (a) 0mm, (b) 5mm, (c) 10mm, (d) 20mm.

### 6.4.2 Cylinder Parallel to $B_0$

As with the sphere, the gradient-echo image of the cylinder shows a loss of intensity at the ends of the tube. However, in the spin-echo image there is no evidence of $T_2$ shortening and the shape of the cylinder remains as expected (Figure 6-11).

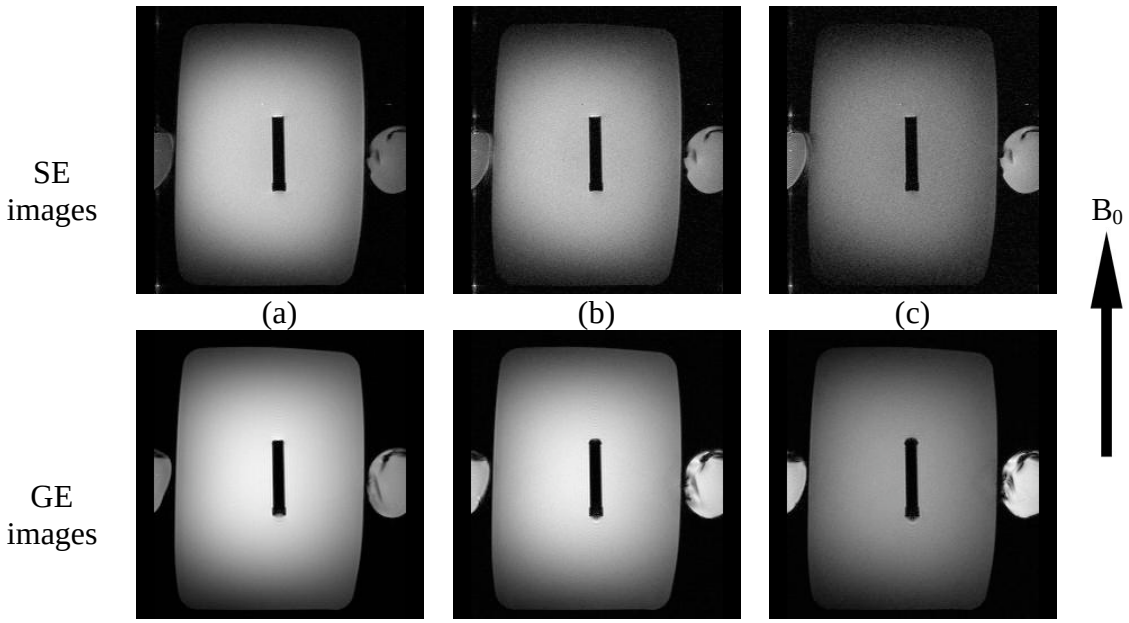



Figure 6-11 – Raw images of central slice for the air-filled cylinder, aligned parallel to the field direction.

(a) TE = 6ms, (b) TE = 12ms, (c) TE = 18ms. The top row of images were obtained from the spin-echo scan, the images on the bottom row were obtained from the gradient-echo scan. The static-field direction is indicated by the arrow.

The $R_2'$ maps of the cylinder are depicted in Figure 6-12. Five slices are shown in the figure – the image from the slice that cuts through the largest cross section of the tube, the two slices directly above this slice and the two slices below. These images show a distinct $R_2^*$ enhancement at the ends of the test tube but there is no evidence of these effects at the sides of the tube. This enhancement is most intense in the central slice and decreases as the slices move away from the centre. These images also show that, as with the sphere, there is still some $R_2^*$ enhancement evident in the slices where there is no tube present.

Figure 6-13 shows how the maximum $R_2'$ (located in the hyper-intense lobes at the ends of the cylinder) decays from the central slice outwards. The largest $R_2'$ is $159.1s^{-1}$ and occurs in the centre slice. The measured $R_2'$ decays away as the imaged slice is moved away from the centre. In the slices immediately outside the dimensions of the tube show a maximum enhancement of $9 – 11s^{-1}$.

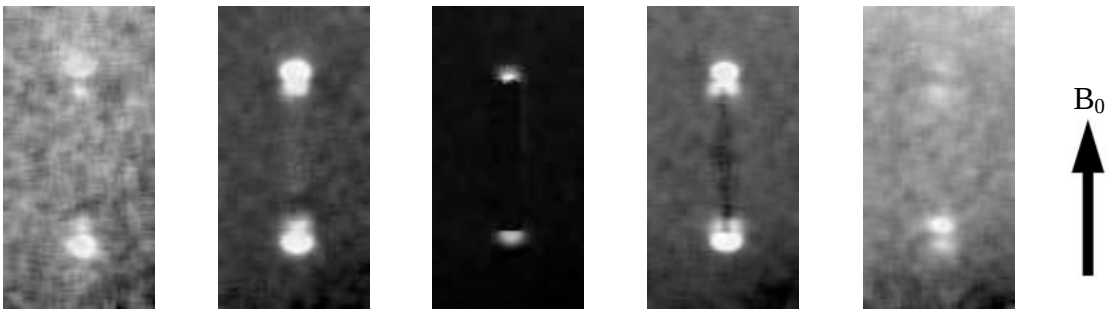



Figure 6-12 – $R_2'$ maps for the air-filled cylinder.

(a) 20mm below the centre slice, (b) 10mm below the centre slice, (c) in the centre slice, (d) 10mm above the centre slice, (e) 10mm above the centre slice. The static-field direction is indicated by the arrow.

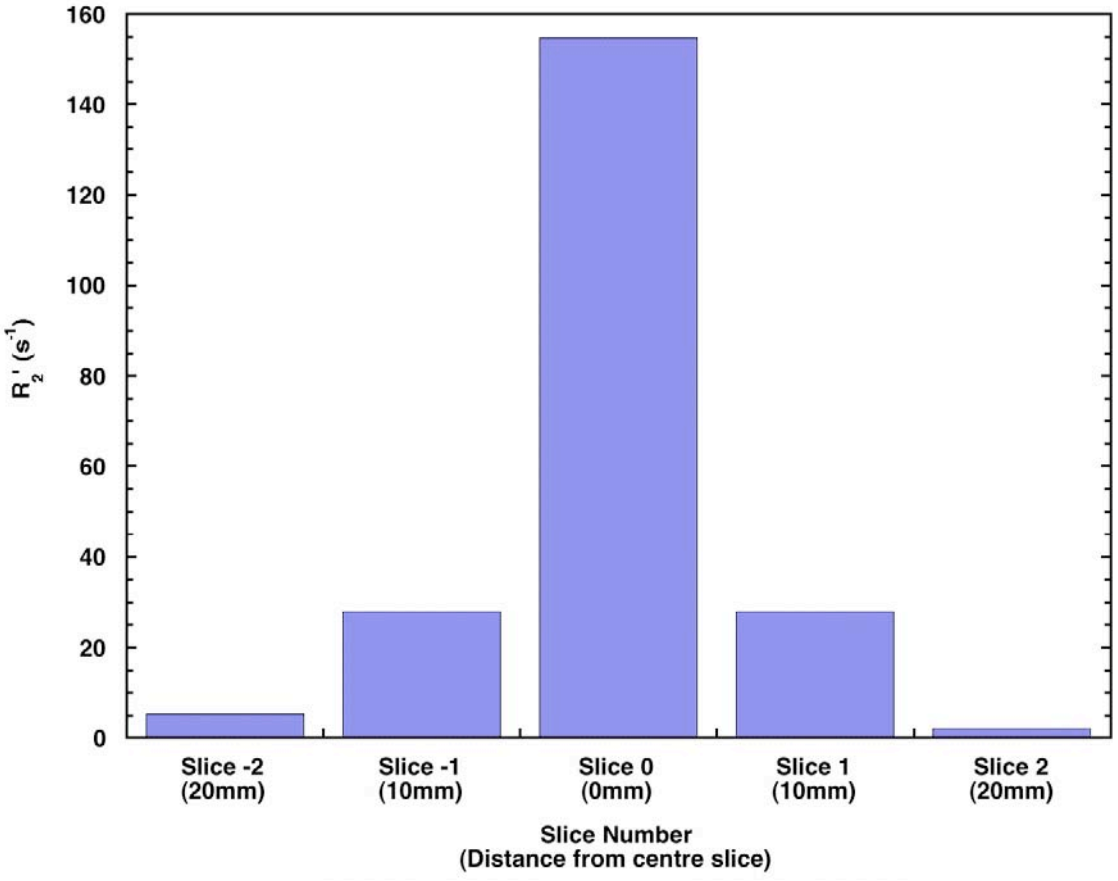



Figure 6-13 – The $R_2'$ measured at the point at the "north pole" of the cylinder for the central slice as well as slices above and below the central slice.

### 6.4.3 Cylinder Orthogonal to $B_0$

The raw spin-echo and gradient-echo images for the cylinder aligned orthogonal to the applied filed direction are in Figure 6-14. With this orientation, images were obtained with the phase encoding direction either parallel or orthogonal to the applied field

direction. $R_2'$ maps were calculated from the $R_2$ and $R_2$* maps for the pixels in the vicinity of the tube Figure 6-15. For both phase encoding directions, the mean $R_2$ (±SD) was 102.6 (± 2.7)$s^{-1}$ and the mean $R_2$* (not including the regions affected by susceptibility artefact) was 101.7 (± 2.7)$s^{-1}$ (phase encoding in the direction of $B_0$) and 103.2 (± 2.6)$s^{-1}$ (phase encoding anterior-posterior). These relaxation rates are not significantly different within the quoted errors and hence there is a zero $R_2$' outside the region affected by the susceptibility interface.

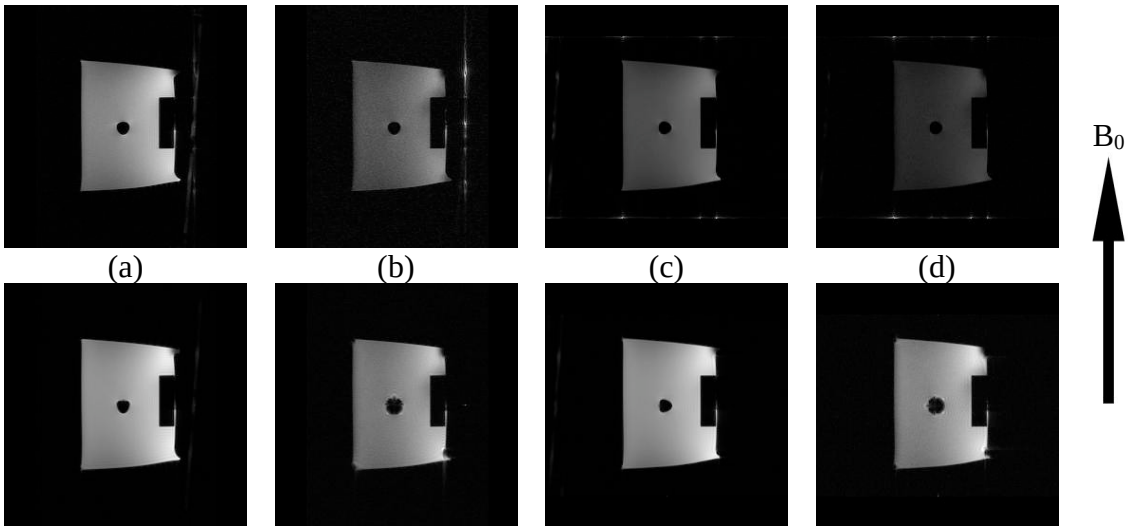



Figure 6-14 – Raw images of central slice for the air-filled cylinder, aligned orthogonal to the magnetic field direction.

(a) TE = 6ms (phase encoding direction parallel to $B_0$), (b) TE = 18ms (phase encoding direction parallel to $B_0$), (c) TE = 6ms (phase encoding direction perpendicular to $B_0$), (c) TE = 18ms (phase encoding direction perpendicular to $B_0$). The top row of images were obtained from the spin-echo scan, the images on the bottom row were obtained from the gradient-echo scan. The static-field direction is indicated by the arrow.

In both cases, there are regions of high and low $R_2'$ alternating around the circumference of the cylinder cross-section (Figure 6-15). The patterns of signal enhancement and subtraction are the same in both phase encoding orientations with a 90º rotation. Signal enhancement due to $R_2'$ effects was up to 30$s^{-1}$ in the pixels adjacent to the cylinder cross section. The low signal regions in the images had $R_2$* as low as 20$s^{-1}$ below the mean $R_2$ seen in the Mn-doped gelatine.

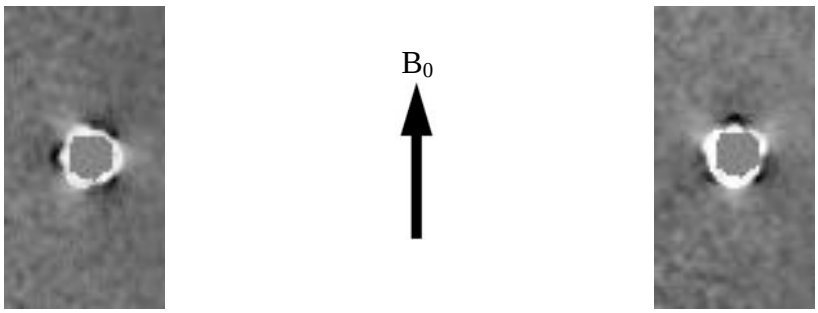



Figure 6-15 – $R_2'$ maps for the cylinder aligned orthogonal to the magnetic field direction with different phase encoding directions.

(a) phase encoding direction parallel to $B_0$, (b) phase encoding direction perpendicular to $B_0$.

## 6.5 Discussion

### 6.5.1 Image Distortion

In both the spin-echo and gradient-echo images of the sphere, some distortion of the circular shape expected for the sphere cross-section is observed. This effect exists in both image sets and is independent of echo time selection. Although it appears more pronounced in the gradient-echo images, this may be a result of some $T_2$* shortening due to the susceptibility artefacts in the vicinity of the sphere. Theory from echo-planar imaging (EPI) predicts [161] that the additional field gradients created by the hollow sphere can cause distortions in the image, owing to in-plane perturbations to the magnetic field.

The results from the orthogonal cylinder experiment support this theory. The $R_2'$ field for this geometry does not display the four-lobe pattern predicted from the theory. In fact, the six-lobe pattern observed alternates between $R_2$* enhancement and reduction. Theoretically the effect of gradients will only cause an increase in the observed relaxation rate in the vicinity of the magnetic susceptibility interface, suggesting a secondary effect is contributing to the signal loss. The magnetic susceptibility effects displayed in this study will not only affect the relaxation of the protons, but can also manifest in altering the phase or frequency encoding of the protons. This change in phase or frequency, will cause the signal to appear in a different part of k-space. When measuring this signal in image space, some of the signal from a given spatial location will appear in another voxel, causing distortions from the expected geometry.

### 6.5.2 $R_2$* Enhancement at Magnetic Susceptibility Interfaces

Aside from the spatial distortions, severe $T_2$* shortening at the surface of the ball is observed in the gradient-echo images. $T_2$* shortening is especially apparent at the poles of the sphere that are aligned with the applied field direction. Compared with the medium, the air inside the sphere has a less negative magnetic susceptibility and hence appears as a dipole of paramagnetic material in a relatively diamagnetic surrounding. In the case of a sphere, the field generated will extend into the medium and behave as a magnetic dipole outside the surface of the sphere. These additional field gradients in the direction of the applied magnetic field (i.e. the z-direction) will result in a difference in precession frequencies across each voxel, causing a rapid dephasing and subsequent shortening of $T_2$*. In comparison, this effect is not observed in the spin-echo images because the loss of signal due to field strength differences on distances larger than the diffusion length are recoverable.

The fact that the severe signal loss is only seen in the gradient-echo images and not in the spin-echo images suggests that this artefact is due entirely to $T_2'$ effects. The relaxometry performed on the phantoms and the subsequent $R_2'$ maps illustrate the magnitude and extent of the susceptibility artefacts. In the case of the spherical objects, these effects are observed at the poles of the sphere and also around the equatorial cross section of the sphere. Little or no $R_2$* enhancement was observed between these regions, indicating that there is no component of the field gradient contributing to additional relaxation of the protons in this region. Qualitatively, this agrees very well with the theory. Hyper-intense regions of $R_2'$ were also observed at the poles for the cylinder aligned parallel to $B_0$, but unlike the sphere, no enhancement was observed away from this region. Along the long edge of the cylinder, the field gradients are small, so any additional relaxation owing to field gradients in these voxels is negligible. Although the case of a finite cylinder cannot be solved analytically, these observations agree with what would be expected conceptually.

### 6.5.3 Agreement Between Experiment and Theory

It is evident from the scaled $R_2'$ maps that the largest $R_2'$ enhancement in the sphere is offset from the equator and occurs in the slices either side of the equator. This property is most likely due to the curvature of the sphere that exists within the depth of a given voxel. For example, in the equatorial slice, the field gradient experienced by each proton in a voxel will not vary much throughout the voxel due to the small variation in

cross-sectional diameter of the sphere at the centre of the slice and the edges. However, in the adjacent slices, the diameter of the sphere cross-section will vary much more across the extent of the slice, resulting in a much larger difference in the relaxation properties within a given voxel, leading to a much quicker dephasing. This off-centre intensity peak is predicted by the theory (Figure 6-10) for a slice of non-zero thickness, supporting the fact that this maximum is due to relative curvature throughout the thickness of a single slice. The theory predicts however, that for larger slice thicknesses, the location of the maximum $R_2'$ enhancement will exist in the equatorial slice. For a 4cm diameter sphere, a slice thickness of larger than 1.5cm would mean that the slices adjacent to the equatorial slice will include the sphere as well as the medium above the sphere, introducing partial volume effects and reducing the measured $R_2^*$ enhancement. The theory also predicts a minimum signal enhancement to be observed around 18mm from the central slice. This minimum may not be seen in the experimental data because it occurs between the 15mm and 22.5mm slice centres. Closer spaced and thinner slices will allow for the observation of this expected region of low $R_2'$.

For the analysis shown for the sphere, the theory and experimental results showed a complex relationship between the maximum observed $R_2'$ and the position of the slice. The curvature of the ball was shown to further enhance $R_2'$ in slices with a non-zero slice thickness. In the case of a cylinder parallel to $B_0$, the ends are flat, so no additional dephasing throughout a voxel due to differences in curvature will exist. Hence, the maximum intensity for the cylindrical geometry occurs in the slice at the centre of the cylinder, which was observed in the experiment.

The theoretical description of the field gradients shows very good agreement with the $R_2'$ maps shown for the hollow sphere. It is at the fourth slice above the equator (or at about 1cm above the sphere) that the lobes at the north and south poles start to drop within the noise level. However, the band across the equator of the ball is still quite intense even at this distance from the ball. The enhancement at the poles of the sphere in some slices extends to up to 20mm into the medium. The use of *in vivo* relaxometry analyses involving $R_2^*$ measurements are sensitive to these variations well beyond the susceptibility interface, potentially causing undesired artefacts in the region of interest. In fact, the signal observed in a given region of interest may be adversely affected by objects or susceptibility interfaces that are outside the field of view of the image. Biomedical studies could involve air-filled cavities in tissue or iron-loaded tissue, where

iron concentrations can be high enough to reduce tissue magnetic susceptibility to zero or even result in a positive magnetic susceptibility [71].

### 6.5.4 Dependence of Magnetic Susceptibility Artefacts on Geometry

It should be noted though that such effects depend highly on geometry and arrangement of objects with varying magnetic susceptibility. In the case of a dipole field, as demonstrated with the sphere, if measurements are being carried out on tissues that fall in the regions between the lobes, then $R_2'$ effects will be minimised. Conversely, if the region of interest lies within a region subject to these artefacts, the $R_2'$ effects will be maximised. This effect has been observed previously [162] where the location of the head relative to the lungs had an effect on the existence of magnetic susceptibility artefacts complicating brain imaging. The results from Truong et al [162] suggest that these effects do have influence over large distances. The theory shown predicts also that the distance at which these effects are significant depends on the size of the object (Figure 6-3). For a sphere of a small radius, the gradient drops to effectively zero at a short distance from the surface of the object. However, the $R_2^*$ enhancement will be larger in the close vicinity of an object with a small effective radius than at a similar distance from a larger object. Conversely, the magnitude of an object of a larger radius will still have significant influence over a much larger distance. This observation supports the findings of Truong et al [162] in that the lungs, a relatively large object filled with air, can have an effect on the brain several centimetres away.

### 6.5.5 Application to the Measurement of Cardiac $R_2^*$

In the context of iron quantification, the location and orientation of an organ such as the lung in relation to the liver or myocardium could cause susceptibility effects in these organs of interest, masking the relaxation due to the iron. Additionally, at the interface of the heart and the liver, the liver has a curvature such that the effective radius of the tissue is large. In a severely iron-loaded liver, it has been shown that sufficiently high iron concentrations can see the net magnetic susceptibility of the liver approach zero and potentially become paramagnetic [71]. So with suitable arrangement of the organs and geometry of the field gradients, the effect from the liver could potentially create artefacts in the cardiac wall.

Current methods of quantifying cardiac iron concentrations specify that the mid-ventricular septum should be used as the region of interest [95] to avoid artefacts. It has

been documented that the cardiac iron concentration is not homogeneously distributed throughout the myocardium [103, 116], so only choosing a small region of interest may not be representative of the true iron concentration. The fact that the theoretical models agree well with the theory means that these models can potentially be used to predict the artefacts due to susceptibility interfaces provided the geometry is known. Additionally, accurate modelling could be used to correct for the artefacts. For example, corrections could be made for $T_2$* shortening due to cardiac veins observed in the myocardium. The corrected images could then be used in relaxometry methods to quantify iron concentrations in the whole myocardium rather than solely in the septum, providing a more accurate indication of total iron.

## 6.6 Summary

The effect of magnetic susceptibility interfaces on relaxometry methods for monitoring body iron stores has been quantified in gelatine phantoms with distinct susceptibility interfaces. Some geometric distortion was evident in raw spin-echo and gradient-echo images as a result of the field gradients perturbing the phase and frequency encoding used to create the images. These field gradients also caused severe $R_2$* enhancement in the vicinity of the interface and these effects were observed to extend well into the surrounding medium, in some cases extending up to 2cm beyond the interface itself. The observed effects were found to agree very well with effects predicted from theoretical analysis of the field gradients in the phantom. These effects are significant and need to be considered carefully when gradient-echo relaxometry methods are used quantitatively as in the quantification of body iron.

# Chapter 7

# The Effect of Analysis Methods on the Measurement of Cardiac $R_2$* and the Assessment of Magnetic Susceptibility Artefacts in Cardiac Relaxometry

## *7.1 Introduction*

Patients with thalassaemia major develop secondary iron overload owing to the regular blood transfusions required for treatment. To remove the excess iron, patients are concurrently treated with iron chelation therapy. Without chelation therapy, excess iron is deposited in parenchymal tissue of organs such as the liver, spleen and also in the heart. Deposition of excessive amounts iron in the myocardium is thought to be responsible for cardiac disease, the most common cause of death in thalassaemia [102, 163]. Cardiomyopathies associated with transfusion related iron deposition have also been documented in non-thalassaemic conditions [103]. Although endomyocardial biopsy is the most direct method of monitoring cardiac iron burden, this procedure has associated risks for the patient and is subject to sampling error [103, 116]. These problems make endomyocardial biopsy inappropriate for applications such as drug trials, as it cannot be used in follow-up studies. Hence, a method for non-invasive quantification of myocardial iron is required to test the efficacy of any drug therapy aimed at reducing cardiac tissue iron deposits, to monitor patients in follow up studies and to effectively manage any drug treatment.

Magnetic resonance imaging (MRI) is currently the only method available to monitor cardiac iron concentrations non-invasively [96]. To date only a few studies have been carried out, each with individual techniques and each adopting differing analysis methods. The importance of the methods used to quantify relaxation rates in tissue has been demonstrated in the liver. No correlation between $R_2$ and liver iron concentration was observed using echo times of 30 – 90ms [164] but significant correlations have been observed with the use of shorter echo times [7, 97, 165]. Different treatments of background noise have also been reported to affect observed relationships between liver iron concentration and relaxation rates [7, 93]. The importance of echo time choice and treatments of signal offsets has been reported recently by Ghugre et al. [166]. Some

techniques for quantifying cardiac iron concentrations are being used clinically and variations in analysis methods will pose a problem if comparisons with results from multiple centres are required.

Owing to the risks associated with endomyocardial biopsy and the inhomogeneous nature of cardiac iron distributions [116], most studies have not involved calibration of observations from MRI with biopsy data. Many of the studies have made comparisons with other body iron measures such as liver iron concentration [120] or serum measures [100, 120, 122]. Anderson et al [95] saw a relationship between cardiac $T_2$* and measures of cardiac function but did not observe a significant trend with liver iron concentration. One recent exception is in the work of Mavrogeni et al [101], which demonstrates the relationship between cardiac $T_2$ and stainable iron from cardiac biopsy specimens. With no calibration with biopsy data, caution must be exercised in interpreting relaxation rates as being indicative of cardiac iron concentrations. Furthermore, $T_2$* values are not only affected by the presence of iron in tissue but also by macroscopic magnetic field inhomogeneities. These inhomogeneities arise from variations in the static field as well as differences in the susceptibility of different tissues. One example is the effect of blood vessels on $T_2$* [130, 131], which generate $R_2$* enhancement in tissue not related to the presence of iron.

This chapter addresses the differences between a variety of analysis techniques to measure cardiac $R_2$*, which currently appears to be the most commonly used parameter for monitoring cardiac iron concentrations. The effects of echo time selection, the choice of mapping the spatial variation of $R_2$* or the use of single measurement region of interest (ROI) techniques and the type of signal decay model to be used in the analysis of the raw MR data will be investigated to assess how these affect the $R_2$* values obtained. The knowledge gained from this investigation will be used to identify the best method for assessing the artefacts in cardiac $R_2$* measurements and to ensure that the method gives similar results to those methods that are used clinically. The appropriate method will be used to assess the effect that imaging artefacts can have on the accuracy and feasibility of cardiac $R_2$* measurements as an indicator of cardiac iron concentration. The relationship between cardiac $R_2$* measurements and liver iron concentration measurements will also be discussed.

## 7.2 Materials and Methods

### 7.2.1 Subjects

Measurements were obtained from 30 adults aged 19 to 51 years (mean 32 years, standard deviation 7 years). Their primary conditions were β-thalassaemia major (n = 22), β-thalassaemia intermedia (n = 4), β-thalassaemia/Haemoglobin E (n = 1), and Diamond-Blackfan anaemia (n = 1). Two healthy volunteers who had never received a blood transfusion were used as controls. Of the iron loaded subjects, all except one were managed with regular transfusions of packed red cells and conventional iron chelation therapy (subcutaneous desferrioxamine). Therapy had been directed with regular serum ferritin measurements and compliance with chelation plans was variable.

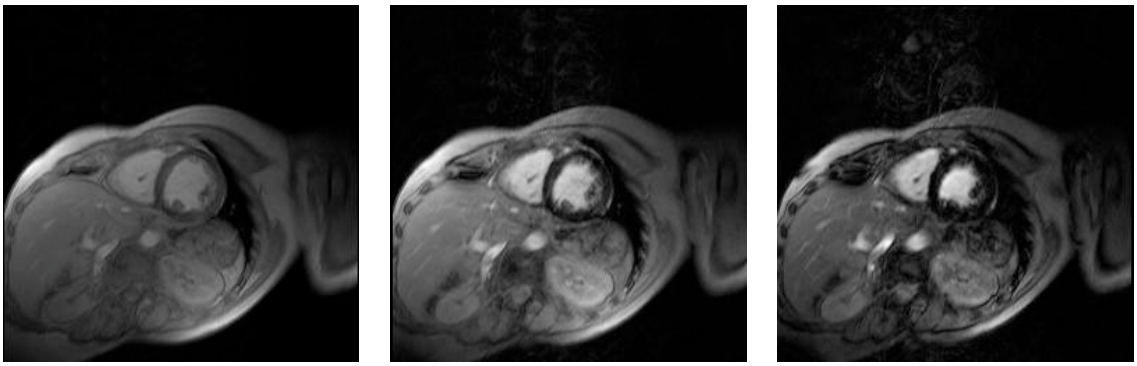
### 7.2.2 Magnetic Resonance Image Acquisition

Subjects were examined on a 1.5 Tesla cardiac scanner (Siemens Sonata). Two separate examinations were performed on each subject. Firstly, a spin-echo sequence was used to obtain a series of 19 single spin-echo axial liver images (TR = 2500ms, TE = 6, 7, 8, 9, 12, 15, 18ms). The field of view (FOV) and matrix size were allowed to vary according to the patient size. A slice thickness of 5mm and inter-slice spacing of 5mm was used for all patients. Of the 19 slices obtained, the slice with the largest cross-sectional area of liver tissue was used in subsequent analyses.

The subjects were examined again with 11 echocardiograph (ECG) triggered segmented gradient recalled echo (GRE) sequences in the cardiac short-axis oblique with imaging parameters TR = heart rate, TE = 3.6, 4.6, 5.6, 6.6, 7.6, 8.6, 9.6, 12, 14, 16, 18ms, flip angle = 35°, FOV 350 x 350 mm, Matrix 256 x 144, slice thickness 10mm (Figure 7-1). The images ran through the middle portion of the ventricular septum and included cross sections of the liver and left ventricular cardiac tissue. The segment acquisition period was fixed for all echoes at 186 ms. Acquisition occurred at each R-wave (delay time = 0ms). The majority of patients (14) had both scans performed on the same day. The longest time elapsed between the two scans was 16 days.

After image acquisition, the cardiac short-axis oblique images were registered manually. Only pixel-wise translational registration was performed on this set of images. Interpolation algorithms required to perform rotational registration were found to adversely affect the background noise profile in the images. As modelling the

background noise is an integral part of the data analysis technique used [142], rotational registration was avoided. The images were registered based on the location and orientation of the cardiac septum. Through-plane misregistration was also observed in some cases but this was not corrected.



(a) (b) (c)

Figure 7-1 – Sample gradient-echo images including cross-sections of the liver and left ventricular myocardium.

(a) TE = 3.6ms, (b) TE = 9.6ms, (c) TE = 18ms.

### 7.2.3 MRI Analysis Methods

#### Liver

The images of the liver obtained from the spin-echo scans were processed and analysed as in Chapter 2. The same analysis method was also applied to the series of gradient-echo images, which were used to create $R_2$* maps of the liver. Owing to the orientation of the slices in the gradient-echo images, spin-density projection and gain correction were not performed. Of the 30 subjects, liver $R_2$* maps were only generated for 21 subjects as the signal dropped below the noise floor of the image at very early echo times for the remaining 9 subjects owing to high liver iron concentrations, thus corrupting the exponential decay profile.

#### Heart

In a similar fashion to the analyses for the liver in the gradient-echo images, the analyses of the myocardium did not involve spin-density projection or gain correction. Owing to the thickness of the myocardium, the raw images were processed with a 3x3 pixel smoothing window prior to analysis rather than the 5 x 5 window used in the liver. Septal $R_2$* values were quantified using eight different methods to process the raw gradient-echo images obtained from the segmented gradient echo sequence. Of these

eight methods, four utilised region of interest (ROI) methods while the remainder used pixel-wise mapping techniques. A maximum of only eight spin-echo images could be included in the pixel-wise mapping analyses owing to limitations in the software used to generate the $R_2$* maps. All mapping analyses were performed by the same analyst with a second analyst performing the ROI analyses. While the majority of the methods involved fitting a single exponential function to the signal intensity against echo time, two of the methods involved fitting a bi-exponential decay curve to the intensity values. In each method a subset of at least 5 of the 11 echo times was used in the analysis for each subject. An additional test was performed on two of the ROI methods, where the ROI at each echo time was required to satisfy the condition:

$$Mean_{ROI} - SD_{ROI} > Mean_{Noise} + SD_{Noise}$$

Where $Mean_{ROI}$ and $SD_{ROI}$ are the mean and standard deviation of signal intensity in the ROI and $Mean_{Noise} + SD_{Noise}$ are the mean and standard deviation of signal intensity in a region of free space in the image. The echo times that failed this test were excluded from the final analysis. The details of all eight methods are summarised in Table 7-1.

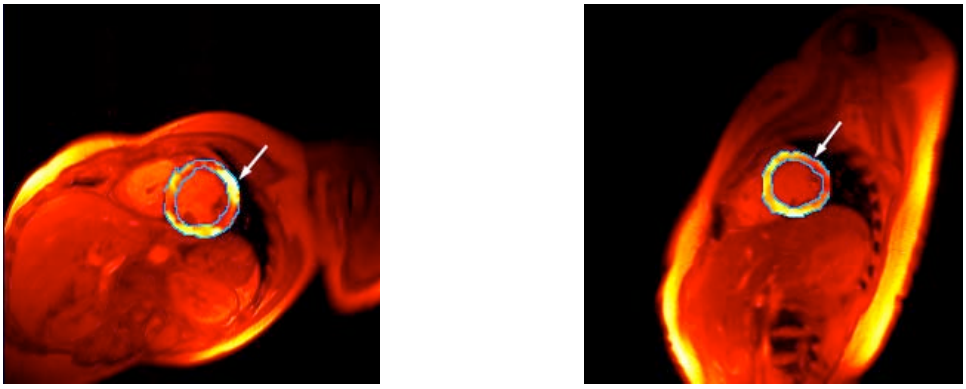
| Method | Echo times (ms) | ROI / Map | Decay Model | Signal test |
|---|---|---|---|---|
| 1 | 3.6, 5.6, 7.6, 9.6, 12, 14, 16, 18 | Map | Bi | No |
| 2 | 3.6, 5.6, 7.6, 9.6, 12, 14, 16, 18 | Map | Mono | No |
| 3 | 3.6, 4.6, 5.6, 6.6, 7.6, 8.6, 9.6 | Map | Bi | No |
| 4 | 3.6, 4.6, 5.6, 6.6, 7.6, 8.6, 9.6 | Map | Mono | No |
| 5 | 5.6, 6.6, 7.6, 8.6, 9.6, 12, 14, 16, 18 | ROI | Mono | No |
| 6 | 3.6, 4.6, 5.6, 6.6, 7.6, 8.6, 9.6, 12, 14, 16, 18 | ROI | Mono | No |
| 7 | Subset from Method 5 that satisfied the signal test | ROI | Mono | Yes |
| 8 | Subset from Method 6 that satisfied the signal test | ROI | Mono | Yes |

Table 7-1 – Description of each of the eight methods used to quantify septum $R_2$*.

For five of the subjects, images could not be corrected for misregistration issues owing to the severity of the misregistration thus giving unreliable results for the mapping methods. As a result, these subjects (four β-thalassaemia major subjects and one β-thalassaemia intermedia subject) were excluded from the subsequent analyses. This problem could possibly have been avoided by a single breath hold acquisition similar to that used previously [167].

### 7.2.4 Artefact Analysis

To assess the effect of artefacts in $R_2$* maps of cardiac tissue, a region of interest comprising the spatial extent of the artefact in the vicinity of the great cardiac vein (GCV) was chosen by eye from the $R_2$* map generated by method 1 for each subject. This hyper-intense region of interest in the $R_2$* image included all pixels which were judged to be representative of the artefact and not of the cardiac tissue. In the cases where no such hyper-intense region was observed, it was assumed that the pixels in that region of the myocardium were not affected by artefacts and were indicative of the physical properties of the myocardium. Figure 7-2 shows two examples of left ventricular $R_2$* maps. While the hyper-intense artefact was observed in most of the subjects (Figure 7-2a), it was not observed in others (Figure 7-2b). Another artefact was observed near the posterior vein of the left ventricle in 15/27 subjects. This artefact had a more variable position, size and a lower relative intensity than the artefact near the GCV. The magnitude of this artefact was not quantified in this work.



(a) (b)

Figure 7-2 – Sample septum $R_2$* maps superimposed on false colour gradient echo images.

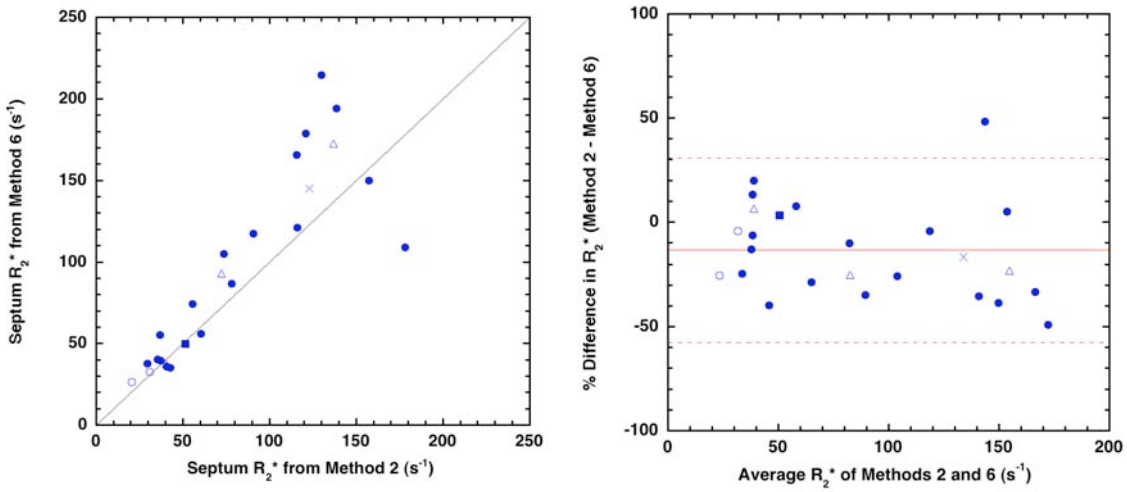
(a) Artefact present in the vicinity of the great cardiac vein (indicated by the arrow), (b) No artefact present in the vicinity of the great cardiac vein (arrow).

## 7.3 Results

### 7.3.1 Comparison between methods of measuring heart $R_2$*

#### Mapping vs ROI methods

Septum $R_2$* as measured by method 6 and method 2 (Figure 7-3a) are strongly correlated (Spearman's rank order test - $\rho = 0.90$, $p < 0.0001$). The difference in the mean $R_2$* value for the septum, measured using methods 2 and 6, was determined as a percentage of the average of the two $R_2$* values for each subject. Spearman's rank order test shows no significant evidence of a correlation ($p = 0.12$) between the percentage difference and the mean $R_2$*, as measured from these two methods (Figure 7-3b). The mean percentage difference of -13% is significantly different from zero ($p = 0.004$). The 95% limits of agreement for these data (± the 95% confidence limits) are –58 (±15) % to 31 (±15) %.




(a) (b)

Figure 7-3 – Comparison of methods 2 and 6.

(a) $R_2$* as measured by method 6 plotted against $R_2$* as measured by method 2, (b) Bland-Altman plot. β-thalassaemia major (●), β-thalassaemia intermedia (△), β-thalassaemia/haemoglobin-E (×), Diamond-Blackfan anaemia (■), controls (○).

#### Comparing decay models

Septum $R_2$* as measured by method 1 and method 2 (Figure 7-4a) are strongly correlated (Spearman's rank order test - $\rho = 0.98$, $p < 0.0001$). The difference in the mean $R_2$* value for the septum, measured using methods 1 and 2, was determined as a percentage of the average of the two $R_2$* values for each subject. Spearman's rank order

test shows a statistically significant correlation (ρ = 0.61, p = 0.001) between the percentage difference and the mean $R_2$*, as measured from these two methods (Figure 7-4b). The mean percentage difference of 18% is significantly different from zero (p < 0.0001). The 95% limits of agreement for these data (± the 95% confidence limits) are –13 (±11) % to 49 (±15) %. The relationship between the septum $R_2$* as determined by method 1 and method 2 was found to be non-linear as determined from a log-log plot of the two methods (p < 0.0001).

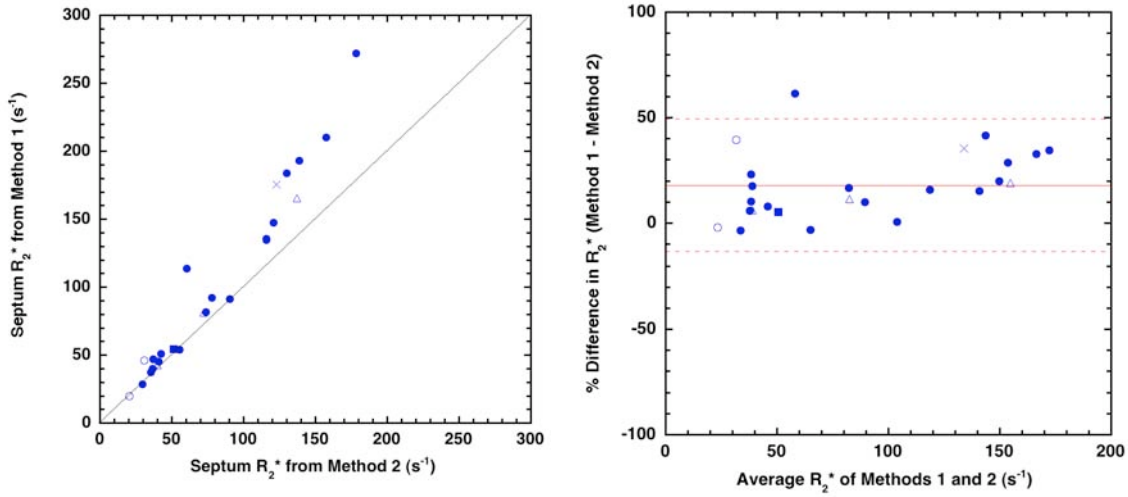



Figure 7-4 – Comparison of methods 1 and 2.

(a) $R_2$* as measured by method 1 plotted against $R_2$* as measured by method 2, (b) Bland-Altman plot. β-thalassaemia major (●), β-thalassaemia intermedia (△), β-thalassaemia/haemoglobin-E (×), Diamond-Blackfan anaemia (■), controls (○).

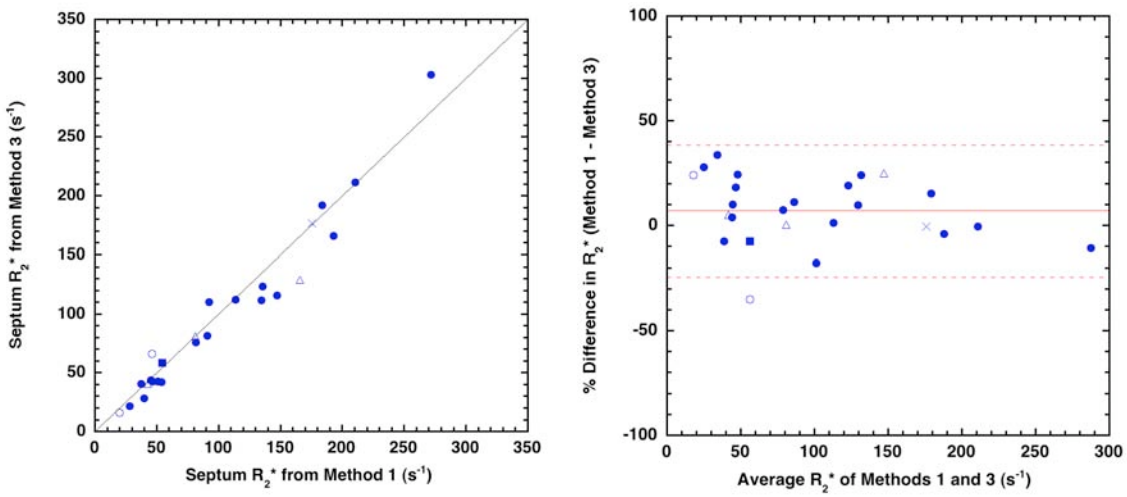
## Effect of echo times used in analysis

Septum $R_2$* as measured by method 1 and method 3 (Figure 7-5a) are strongly correlated (Spearman's rank order test - ρ = 0.98, p < 0.0001). The difference in the mean $R_2$* value for the septum, measured using methods 1 and 3, was determined as a percentage of the average of the two $R_2$* values for each subject. Spearman's rank order test shows no significant evidence of a correlation (p = 0.14) between the percentage difference and the mean $R_2$*, as measured from these two methods (Figure 7-5b). The mean percentage difference of 7% is significantly different from zero (p = 0.02). The 95% limits of agreement for these data (± the 95% confidence limits) are –24 (±12) % to 38 (±12) %.

Septum $R_2$* as measured by method 2 and method 4 (Figure 7-6a) are strongly correlated (Spearman's rank order test - ρ = 0.96, p < 0.0001). The difference in the mean $R_2$* value for the septum, measured using methods 2 and 4, was determined as a

percentage of the average of the two $R_2$* values for each subject. Spearman's rank order test shows a statistically significant correlation ($\rho$ = 0.47, p = 0.019) between the percentage difference and the mean $R_2$*, as measured from these two methods (Figure 7-6b). The mean percentage difference of -2% is not significantly different from zero (p = 0.28). The 95% limits of agreement for these data (± the 95% confidence limits) are –36 (±12) % to 32 (±12) %.

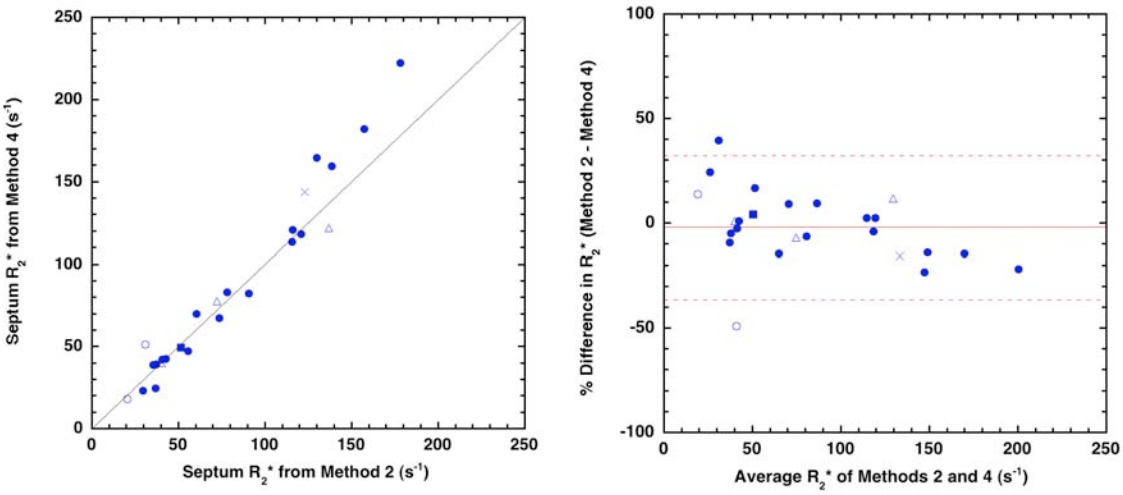
Septum $R_2$* as measured by method 5 and method 6 (Figure 7-7a) are strongly correlated (Spearman's rank order test - $\rho$ = 0.99, p < 0.0001). The difference in the mean $R_2$* value for the septum, measured using methods 5 and 6, was determined as a percentage of the average of the two $R_2$* values for each subject. Spearman's rank order test shows no significant evidence of a correlation (p = 0.91) between the percentage difference and the mean $R_2$*, as measured from these two methods (Figure 7-7b). The mean percentage difference of -1% is not significantly different from zero (p = 0.31). The 95% limits of agreement for these data (± the 95% confidence limits) are –20 (±7) % to 18 (±7) %.




(a) (b)

Figure 7-5 – Comparison of methods 1 and 3.

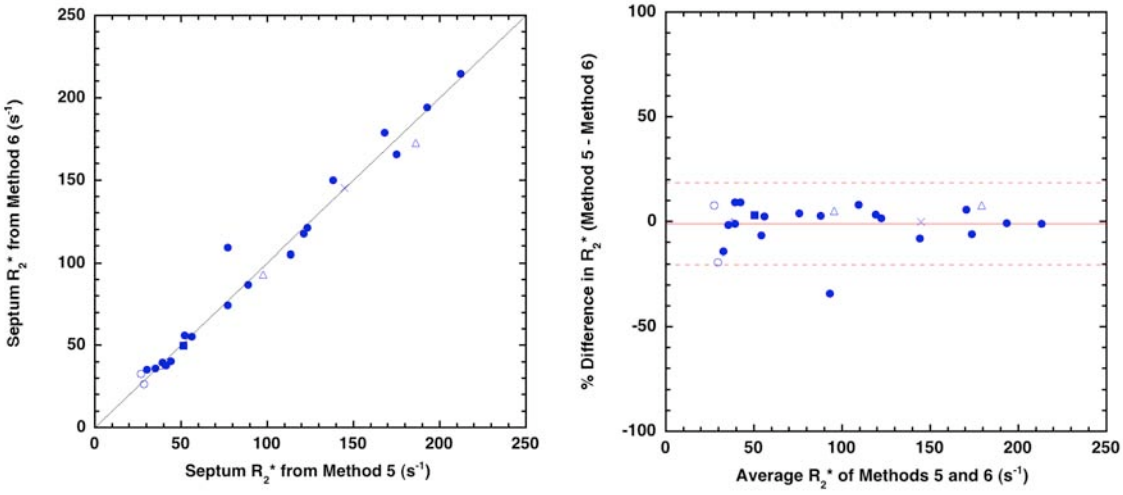
(a) $R_2$* as measured by method 3 plotted against $R_2$* as measured by method 1, (b) Bland-Altman plot. β-thalassaemia major (●), β-thalassaemia intermedia (△), β-thalassaemia/haemoglobin-E (×), Diamond-Blackfan anaemia (■), controls (○).




(a) (b)

Figure 7-6 – Comparison of methods 2 and 4.

(a) $R_2$* as measured by method 4 plotted against $R_2$* as measured by method 2, (b) Bland-Altman plot. β-thalassaemia major (●), β-thalassaemia intermedia (△), β-thalassaemia/haemoglobin-E (×), Diamond-Blackfan anaemia (■), controls (○).



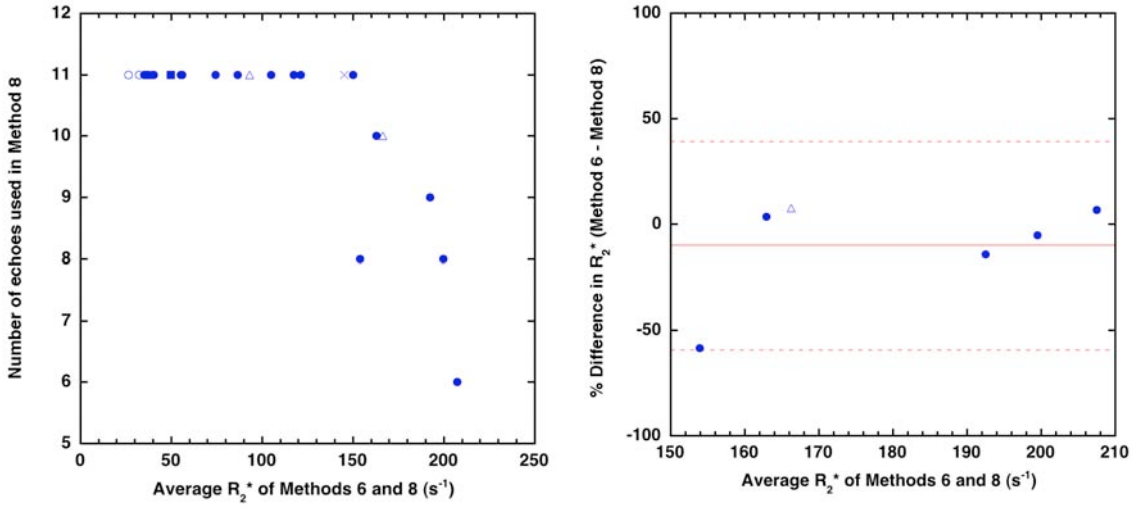

(a) (b)

Figure 7-7 – Comparison of methods 5 and 6.

(a) $R_2$* as measured by method 6 plotted against $R_2$* as measured by method 5, (b) Bland-Altman plot. β-thalassaemia major (●), β-thalassaemia intermedia (△), β-thalassaemia/haemoglobin-E (×), Diamond-Blackfan anaemia (■), controls (○).

### Effect of signal test

When subject to the signal test described, six of the patients required that some of the later echo times be removed to ensure that all echoes remain above the acceptable signal level. The number of echo times removed from the analysis varied between patients with up to five echoes removed in one subject (Figure 7-8a). All six cases requiring the removal of some echo times had an average septum $R_2$* above 150$s^{-1}$. For the subjects with an average septum $R_2$* below 150$s^{-1}$, all 11 echoes were used as they all met the criteria of the signal test. For the six subjects that had some echoes removed, the mean percentage difference (between methods 6 and 8) of -10% is significantly different from zero (p = 0.03). The 95% limits of agreement for these data (± the 95% confidence limits) are –59 (±17) % to 39 (±17) % (Figure 7-8b).




(a) (b)

Figure 7-8 – Comparison of methods 6 and 8.

(a) Number of echo times used in method 8 plotted against the average septum $R_2$* as measured by methods 6 and 8, (b) Bland-Altman plot for the six subjects that required the removal of some echoes to meet the criteria of the signal test. β-thalassaemia major (●), β-thalassaemia intermedia (△), β-thalassaemia/haemoglobin-E (×), Diamond-Blackfan anaemia (■), controls (○).

## 7.3.2 Relationship Between Septum $R_2$* and Liver $R_2$

For each of the patients, the septal $R_2$* was calculated using each of the eight described methods and compared with the mean liver $R_2$, as determined from the single spin-echo images. Figure 7-9 shows how septal $R_2$*, as measured by the method 2, varies with liver $R_2$. For this cohort of patients, a moderate but statistically significant relationship

is observed between these two parameters (Spearman's rank order test ρ = 0.63, p = 0.001).

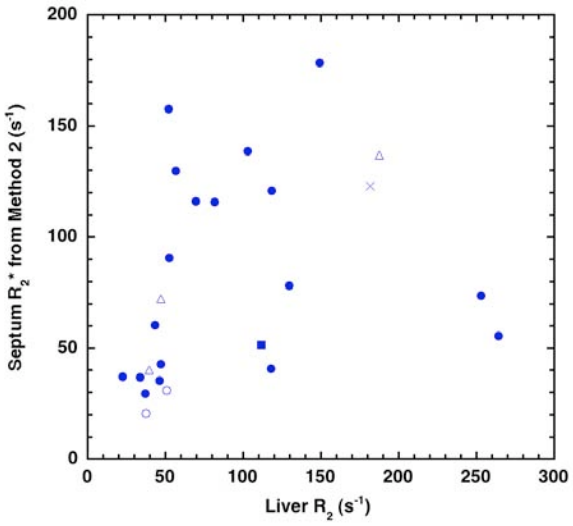



Figure 7-9 – Septum $R_2$* (as measured by method 2) plotted against liver $R_2$.

β-thalassaemia major (●), β-thalassaemia intermedia (△), β-thalassaemia/haemoglobin-E (×), Diamond-Blackfan anaemia (■), controls (○).

The results of Spearman's rank order test for comparisons between liver $R_2$ with all of the methods used to measure septal $R_2$* are shown in Table 7-2. Also in Table 7-2 are the results of Spearman's rank order test when each of the methods are compared to liver $R_2$*. The values shown for liver $R_2$* are on the subset of 21 subjects in which the $R_2$* maps were not corrupted by noise.

A statistically significant relationship is seen for all of the methods when measured against liver $R_2$. The mapping methods using mono-exponential signal decay modelling exhibit the highest ρ values while the ROI methods that used no signal test for the removal of bad echoes showed the lower of the Spearman ρ results. Similar results were obtained in comparisons with the liver $R_2$* data.

However, performing a z-test on all of the methods for both liver $R_2$ and $R_2$* shows that there is no significant difference between any of the correlation coefficients ($p > 0.05$ for all pairs of correlation coefficients).

| | Liver $R_2$ | | Liver $R_2$* | |
|---|---|---|---|---|
| Method | ρ | p | ρ | p |
| Method 1 | 0.575 | 0.003 | 0.584 | 0.009 |
| Method 2 | 0.625 | 0.001 | 0.649 | 0.003 |
| Method 3 | 0.530 | 0.006 | 0.577 | 0.010 |
| Method 4 | 0.585 | 0.002 | 0.591 | 0.008 |
| Method 5 | 0.512 | 0.009 | 0.470 | 0.042 |
| Method 6 | 0.519 | 0.008 | 0.500 | 0.029 |
| Method 7 | 0.539 | 0.006 | 0.544 | 0.016 |
| Method 8 | 0.547 | 0.005 | 0.570 | 0.011 |
| Average | 0.554 ± 0.038 | 0.005 | 0.561 ± 0.056 | 0.016 |

Table 7-2 – Table of Spearman rank order correlations between septum $R_2$* determined from each method and transverse relaxation rates ($R_2$ and $R_2$*) in the liver.

### 7.3.3 Quantitative Assessment of Artefacts

In 18 of the 25 observed subjects a hyper-intense $R_2$* region is observed in the vicinity of the great cardiac vein (GCV). Over all subjects in which this region was observed, the average magnitude of the $R_2$* increase (±SD) resulting from the artefact is about 230 (±74)% (Figure 7-10). All of the subjects in which this artefact was not observed were in the thalassaemia major group.

The $R_2$* enhancement seen at the location of the GCV artefact (the difference between $R_2$* in the GCV and the septum) increases linearly with increasing septum $R_2$* (Figure 7-11). Results from Spearman's rank order test show that this relationship is strong and statistically significant ($\rho = 0.71$, $p = 0.0009$).

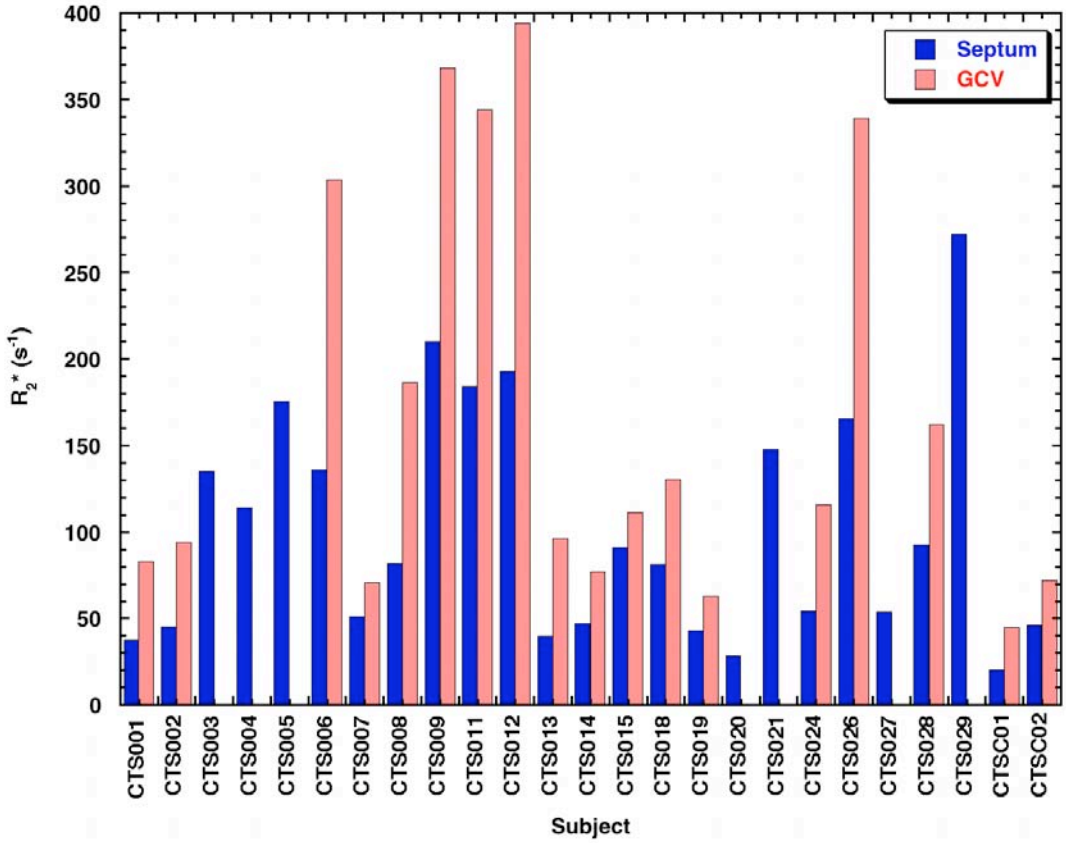



Figure 7-10 – $R_2$* for each patient as measured using method 2 in the septum and in the GCV artefact.

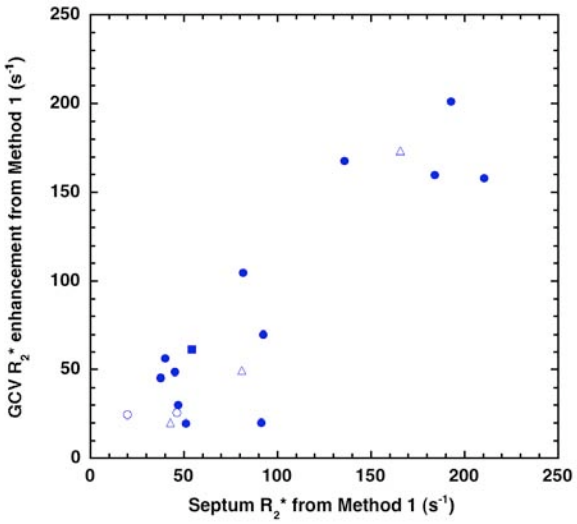



Figure 7-11 – $R_2$* enhancement in the vicinity of the GCV plotted against septum $R_2$*. β-thalassaemia major (●), β-thalassaemia intermedia (△), Diamond-Blackfan anaemia (■), controls (○).

## 7.4 Discussion

### 7.4.1 Variations in Septum $R_2$* Using Different Analysis Methods

The majority of methods investigated to date have used ROI techniques to determine $R_2$* [95, 124]. ROI methods are simpler to perform, as they require less computation time. Mapping techniques however are useful in that they allow one to see the spatial variation of $R_2$* within a given region, although requiring much more attention to detail in the image preparation prior to analysis. Both have their respective advantages and disadvantages but in relaxometry methods, these differences can become important.

These subtle but nevertheless important differences can cause variations in $R_2$* depending on which type of method is used. Compared to all other comparisons in this study, the limits of agreement between the mapping and ROI methods of 45% either side of the mean percentage difference are broad, suggesting a relatively large discrepancy between mapping and ROI methods. One possible reason for this is the fact that different analysts performed the ROI and mapping analyses. The subjective nature of region selection will cause differences in the pixels included in the analysis and hence variations in the $R_2$* value calculated by each method. This issue raises concerns for the reproducibility of $R_2$* when such methods are used in large clinical studies, where multiple centres and multiple analysts may be required. This assumption will need to be verified in subsequent studies. Alternatively, the variations between the $R_2$* values recorded by these two methods could arise from the different treatment of the pixel intensity information by each method in fitting the decay profile.

Echo time selection is of great importance in relaxometry studies. A lack of data points at early echo times, where the most signal decay occurs, can adversely affect the modelling of the decay curve. This problem is particularly relevant for high relaxation rates where the signal can become comparable to the noise at relatively short echo times. Furthermore, an insufficient spread of data points across echo times or the inclusion of data at later echo times where the signal is overly corrupted by noise will also cause problems in fitting an accurate model to the decay curve. Methods 5 and 6 allowed for an investigation into the effects of data points at short echo times. The narrow limits of agreement between these two methods are not surprising owing to the typical rates of decay in iron loaded cardiac tissue [95]. In this cohort of patients, the signal in all cases remained well above the noise floor for all echoes in the majority of

patients. Of the patients in which some of the echoes were corrupted by noise, a sufficient number of images remained where the signal was still dependent on the relaxation properties of the tissue, rather than being affected by noise.

The investigation into the effects of long echo times on the calculated relaxation rates, as was achieved in comparisons of method 2 and method 4, showed that the 95% limits of agreement between these two methods were more widely separated than those for the investigation into the effects of short echo times. The greater variation seen with the later echo times is attributable to the fact that at later echo times, the signal becomes more affected by background noise. The decay rate determined from a fit to the data will vary quite substantially with the inclusion or exclusion of these points as the signal intensity at these echo times may not be entirely related to $R_2$* effects and confounded by noise. For the range of $R_2$* considered in this study, the appropriate inclusion or exclusion of later echo times in the signal decay modelling is more important than having a sufficient number of echo times earlier in the decay curve.

The removal of images at echo times where the SNR is too low is one way to deal with data that is corrupted by noise. This procedure results in a smaller number of data points being used in the signal decay rate modelling, potentially affecting the fit to the decay curve. Removing data points from the exponential signal decay model based on the results of the signal to noise test has no influence for low $R_2$* values as all 11 echo times are included in the signal tested analyses. For $R_2$* values above $150s^{-1}$, a number of later echo times is removed from the decay rate modelling as they do not meet the signal test requirement and a change in measured $R_2$* is observed. This effect is only a concern for the subjects where the signal decays rapidly enough that the SNR becomes so small that the data points at long echo times are largely affected by noise. In the work of Anderson [95], a cut-off $T_2$* value of 20ms ($R_2$* of $50s^{-1}$) was specified where patients were at increased risk of cardiac failure. The fact that the signal test seems to become an issue only at $R_2$* > $150s^{-1}$ ($T_2$* < 6.7ms), much higher than the cut-off, suggests that SNR effects will not affect the classification of patients into low and high risk categories using this criterion. However, other effects such as echo time selection may introduce errors in the measurements of $R_2$* values closer to the cut-off.

In MRI-based hepatic iron concentration measurements, the use of a bi-exponential model has been proposed to distinguish between the different decay rates of the intra-

cellular and extra-cellular iron deposits [8]. The data from this cohort of patients suggests that the bi- exponential model over-estimates that calculated when using mono-exponential decay model. Mono-exponential fits to multi-exponential data will generally underestimate the real decay rate that occurs in the data. Septum $R_2$* as determined from the bi-exponential model increases non-linearly with the $R_2$* calculated from the mono-exponential decay model. The non-linear relationship gives rise to a saturation effect where the mono-exponential method shows a reduced sensitivity for high $R_2$*. This observation suggests that the tissue in the septum is most likely to be multi-compartment in nature and a mono-exponential decay model may not be sufficient to describe the decay rate effectively.

### 7.4.2 Correlation Between Septum $R_2$* and Liver $R_2$

Although the method used to measure septum $R_2$* can give varying results, a statistically significant correlation between the septal and hepatic relaxation rates was found for this cohort of subjects, regardless of which method was used. The results shown here are similar to other research [120, 124] in which a weak but statistically significant relationship was observed. However, Anderson et al. [95] used liver $R_2$* values as an indicator of liver iron concentration and found no significant relationship between the relaxation rates of the two organs.

The correlation observed in the current study between septum $R_2$* and liver $R_2$, suggests that there is some physiological effect on the septum resulting from liver iron loading. There are two likely possibilities. Firstly, as transfusions are required as part of the treatment for these patients, the transfusion may have an effect on heart function [168, 169]. A change in cardiac function would suggest that some characteristics of the myocardial tissue itself must also be altered, which will result in a change in the relaxation properties of the myocardium. Secondly, studies by Anderson [95] have speculated that the changes seen in cardiac $R_2$* are indeed related to cardiac iron concentrations. *In vivo* studies have demonstrated that liver $R_2$ is related to liver iron concentration as measured by liver biopsy [97]. A similar result cannot be verified as easily in the myocardium because of the problems associated with cardiac biopsy. Recent work in animal models has shown a relationship between *ex vivo* cardiac tissue and liver $R_2$* [118] and also a relationship between cardiac $R_2$ and heart iron concentrations [119]. If the septal $R_2$* is indicative of cardiac iron burden, then the correlation with liver $R_2$ would suggest that the iron loading conditions cause iron to

accumulate simultaneously in both hepatic and cardiac tissue. However, owing to the variable chelation history of the patients and the variable efficacy of drug treatments at chelating iron from different tissue, hepatic and cardiac iron concentrations will not necessarily vary together. These confounding factors may partly explain the weakness of the correlation observed between cardiac and hepatic iron measures in human studies.

### 7.4.3 Dependence of Myocardial $R_2$* on Blood Flowing in Micro-vessels

In ROI methods, the analysis of regions outside the septum is not advisable [95] as the analyses can be corrupted by artefacts from the cardiac vessels. Mapping methods, unlike ROI methods, provide a method for quantifying the magnitude and spatial extent of these artefacts. The location of these artefacts are consistent with the major cardiac vessels and have been observed previously [130]. The correlation between the $R_2$* enhancement due to the GCV artefact and the septum $R_2$*, which has also been observed in recent studies [166, 170], suggests that septum $R_2$* may not be solely affected by myocardial iron. If the artefact is related to non-physiological effects, no correlation with septum $R_2$* would be expected. One possibility is that frequent transfusions could put strain on the myocardium, potentially resulting in a change in the size of the GCV but this is unlikely to account for increase in $R_2$* as dilation of GCV would cause a decrease in field gradients in the vicinity of the vessel (Chapter 6). A more likely scenario is that $R_2$* in the tissue is affected by the micro-vessels within the myocardium, since it is the magnetic susceptibility of the blood in the GCV that causes the artefact and this same blood flows through the capillaries of the myocardium.

Shortening of $T_2$* in the myocardium, while observed in transfusion dependent thalassaemia patients, is not seen in sickle cell patients [125]. The $T_2$* shortening could be a result of the different magnetic susceptibility expected for the blood in thalassaemic and sickle cell patients. Ferritin and hemosiderin like iron deposits are found in red blood cells in thalassaemia patients [171] but not observed in sickle cell patients [172]. The high magnetic moment of these deposits is sufficient to cause an increase in the magnetic susceptibility of the blood in circulation. In some cases, these deposits account for up to 50% of the iron in a RBC but only 9% in sickle cell patients [171]. In other anaemias and in normal subjects, this type of iron deposit is not observed and in thalassaemia minor the ratio is less than 1%.

The equations in Chapter 6 show that the field gradients near a macroscopic object are related to the magnetic susceptibility of that object. So in the case of thalassaemia subjects where the magnetic susceptibility of the blood is higher than that of blood in non-thalassaemics, the field gradients around the vessels will be more severe. If the density of these vessels is suitable, an enhancement in $R_2$* is observed [131]. In the myocardium, the capillary density is optimal for $R_2$* enhancement [131].

## 7.5 Summary

Measurements of cardiac $R_2$* provide a non-invasive option to monitor cardiac iron concentrations. Although ROI methods are used clinically, mapping methods could also be used as they give results similar to those found using ROI methods. Mapping methods also provide a way to assess the magnitude and spatial extent of artefacts owing to the major cardiac vessels. The correlation between $R_2$* enhancement in the artefact near the GCV and the $R_2$* of the septum could indicate the dominance of magnetic susceptibility artefacts resulting from the micro-vessels in determining myocardial $R_2$*.

# Chapter 8

# Concluding Remarks

Methods of quantifying proton transverse relaxation rates on a voxel-by-voxel basis provide more information than region of interest (ROI) methods. Average measurements on regions of interest give a description of the mean properties of the tissue without the need for complex computational processes and have some demonstrated clinical value. In ROI methods however, useful information contained in each voxel regarding the properties of a specific spatial location is effectively lost. Voxel-by-voxel mapping provides a method to visualise the spatial variations of relaxation rates in the tissue of interest. Furthermore, these methods allow for the assessment and quantification of the spatial distribution of relaxation rates, which has potential clinical applications.

The work in this thesis has demonstrated two specific applications of voxel-by-voxel mapping of relaxation rates -

1) The distribution and texture of the relaxation rates in the liver is related to the extent of fibrosis in the tissue owing to the iron distribution changes in the presence of fibrous tissue. This relationship can be used to develop a non-invasive predictive model for fibrosis assessment.

2) Assessment of the magnitude and spatial variation of magnetic susceptibility artefacts in $R_2$* quantification and its specific application to the measurement of cardiac iron concentrations.

## *8.1 Texture Analysis of Liver $R_2$ Maps and the Non-invasive Quantification of Liver Fibrosis*

Texture quantification measures by GTSD matrices provide a reliable and appropriate way to assess the spatial variations in liver $R_2$ maps. These methods are effectively an extension of the description of the $R_2$ distributions from histograms such as coefficient

of variation, which are reported to be related to fibrosis staging [97]. A number of measures using this method were related to fibrosis staging.

However, these measures are subject to variations owing to effects other than the stage of fibrosis. For example, many of the parameters considered were dependent on the area of the region of interest used in the analysis. The dependence on area can be minimised, however, by only considering slices that have a large number of pixels, more than 8000, in the ROI. However, the minimising of slice area effects was not sufficient to give a robust measure of texture for some measures as the intra-patient variations observed still dominated the inter-patient variations.

At least a part of the apparently poor performance of these individual texture measures for assessing fibrosis can be attributed to the fact that iron may act as a form of contrast to enhance sensitivity to fibrosis. The discriminatory power of individual texture measures for fibrosis is enhanced in thalassaemia and haemochromatosis patients where the liver iron concentration is substantially elevated. In hepatitis C patients where the iron concentrations were typically lower, no individual texture measure gave a statistically significant distinction between patients with low or minimal fibrosis and those with fibrosis or cirrhosis. When subjects of all pathologies were considered as a whole, a small number of parameters were apparently appropriate but not to the extent seen in only the thalassaemia and haemochromatosis patients. Although individual texture measures in the whole patient set were not sufficient for the prediction of fibrosis, a model combining various measures could potentially improve discrimination.

A candidate model was developed using four of the most appropriate texture measures. While the model was reproducible in a small validation set of subjects, biopsy specimens were not available for these subjects and no conclusive evidence of the validity of the model could be obtained. In the test set of subjects, the model does not perform as well as other tests currently being investigated such as serum tests and other radiographic techniques. The majority of other studies have assessed the respective tests on a homogeneous cohort of patients with only a single specific pathology. The development of a model using this method for only a specific pathology is likely to improve the performance of the model and should be investigated in subsequent studies. For example, owing to the better discrimination of fibrosis seen in the presence of iron-loading, recruiting a larger number of patients with iron overload diseases to develop a

model is likely to give improved discrimination for fibrosis for patients with iron overload. Alternatively, a test such as this may be more appropriate for the discrimination of cirrhosis rather than fibrosis. Other studies utilising the assessment of biochemical markers and other radiographic techniques, have performed better in the discrimination for cirrhosis as compared to the discrimination of fibrosis [53, 58]. A future investigation on a patient set with a sufficiently large number of cirrhotic patients will be useful to assess this assumption.

The lower predictive power of texture measures for fibrosis in the absence of iron could be the result of lower specificity of $R_2$ to iron in patients with low liver iron concentrations. The $R_2$ value measured in a region of tissue is dependent on numerous physical and physiological factors. In the presence of high iron concentrations, the effect of iron on $R_2$ dominates all other mechanisms that affect $R_2$. At iron concentrations less than 0.5 mg/g dry tissue however, $R_2$ is no longer predominantly affected by iron and other factors such as tissue hydration become important. The construction of a predictive model for $R_2$ that includes both iron concentration and tissue hydration gives a better relationship with measured $R_2$ than iron alone. In these conditions, the factors other than iron that affect $R_2$ appear to confound the relationship between $R_2$ and fibrosis.

The algorithms used to process the GTSD matrices, while showing promise for fibrosis assessment, could be further refined by choosing measures more appropriate to texture variations expected in this application. Application of these texture measures to individual components of the bi-exponential decay model may also be a better indicator for fibrosis and should be pursued. Other future work could investigate combining GTSD measures with other techniques in an attempt to improve the performance of these methods further. Measures such as fractal dimension, have been shown to be an indicator of fibrosis in rats [173]. When combined with GTSD texture measures, these may pick up different texture attributes and add to the predictive power for fibrosis. Additionally, texture calculations on other radiographic techniques could provide higher sensitivity than iron to structures resulting from fibrosis. Finally, combinations of the findings from MRI with other radiographic techniques or serum tests may provide a better indicator for the staging of fibrosis.

## 8.2 Assessment of Magnetic Susceptibility Artefacts in Relaxation Rate Maps

The mapping of $R_2$* provides a reliable method of visualising and quantifying the effect of magnetic susceptibility artefacts in relaxometry. Interfaces between regions of different magnetic susceptibility can cause an enhancement in $R_2$* in the vicinity of the interface between the two regions. The difference in magnetic susceptibility can also cause distortions in the shapes of objects in both spin-echo and gradient-echo imaging. Both of these effects result from the magnetic field generated by a given object distorting the magnetic fields used by the MRI unit to generate the images.

In this work, the effects of magnetic susceptibility interfaces were investigated in various geometries in which the magnetic field gradients could be determined analytically. The mapping of relaxation rates in these phantoms provided an opportunity to quantify both the magnitude and spatial extent of the magnetic susceptibility artefacts. The results from the experiments in a spherical model and in a cylinder aligned parallel to the field both showed a good agreement with the results expected from theoretical considerations. When the cylinder was aligned perpendicular to the applied field direction, the experimental results did not correspond with that expected from the theory but could be accounted for by geometrical distortions arising from the effects of the field gradients on spatial encoding.

The artefacts for the geometries investigated were found to extend up to 2cm into the surrounding medium but the magnitude and extent of the effect depends on the effective size and orientation of the object. An object with a smaller effective radius will have a much larger effect in the immediate vicinity of the interface than would be seen for a larger object. However, the region of influence of a magnetic susceptibility artefact is much larger for an object of larger size. The agreement of the results and theory means that mathematical models can be used to quantify the effects of these artefacts in relaxometry experiments for any geometry. For more complex geometries that cannot be solved analytically, the field distortions and effective $R_2$* may need to be assessed using numerical methods. These theoretical considerations may also be useful in correcting for susceptibility artefacts observed in gradient-echo images.

In clinical use, magnetic susceptibility artefacts can potentially complicate gradient-echo relaxometry measurements in the vicinity of an appropriate region of tissue. Such artefacts can occur between air-filled cavities and other tissue, in the vicinity of blood vessels and potentially in the vicinity of a severely iron loaded liver. Interfaces such as these are particularly relevant in the quantification of cardiac iron concentrations through the measurement of cardiac $T_2$*. The mapping methods used to analyse artefacts in synthetic phantoms are applicable to investigating similar effects *in vivo* in the quantification of cardiac iron.

Methods for the measurement of cardiac $T_2$* as used in the quantification of cardiac iron concentrations are based on relaxometry of average intensities in the septum. The results presented here show that the method used to measure relaxation rates can affect the results from relaxometry experiments. Differences in analysis methods such as echo time selection and the treatment of noise in the image were tested in thalassaemia patients and were found to have an influence on the $R_2$* values measured in the cardiac septum. The most notable difference was observed when comparing the use of mono- and bi-exponential fitting models in the analysis of the data, where there was a non-linear relationship between the two models ($p < 0.0001$). This relationship suggests that the cardiac tissue is multi compartment in nature and a mono-exponential model may not be appropriate. These differences however do not give large discrepancies in calculations of the mean $R_2$* in the septum, further evidenced by the fact that the degree of correlation observed between liver $R_2$ and septum $R_2$* is statistically the same in all methods used.

Unlike ROI methods, mapping methods provide information regarding the spatial variations in $R_2$* throughout the myocardium. With maps of $R_2$*, the influence of magnetic susceptibility artefacts can be both visualised and quantified. In this thesis, the artefact in the myocardium adjacent to the great cardiac vein was investigated. A correlation was observed between the $R_2$* enhancement owing to the GCV and the $R_2$* of the septum. This relationship suggests that the septum $R_2$* may not be dominated by tissue iron deposits as previously reported but rather by the magnetic susceptibility of the blood flowing through both the GCV and myocardial capillaries. The higher concentration of ferritin and hemosiderin-like deposits in the blood in thalassaemia patients could be an explanation for this phenomenon. The hypothesis is that the presence of this blood in micro-vessels in the myocardium generates a magnetic field

that acts to enhance the $R_2$* in the myocardium, including in the septum. In the myocardium, the density of the micro-vessels is optimal for the enhancement of $R_2$* [131]. The $R_2$* enhancement observed in the myocardium in thalassaemia is not observed in other anaemias such as sickle cell disease where the ferritin and hemosiderin-like deposits are not observed to the same extent [171, 172]. This hypothesis can be verified by future research in a number of ways. Firstly, research needs to be done to investigate the magnetic susceptibility of the blood in thalassaemia patients and in other anaemias by direct measurement, to quantify the difference. Secondly, the hypothesis could be verified by performing similar analyses of the artefacts near the GCV on non-thalassaemic subjects. A third alternative would be to investigate the effects of perfusing blood through the capillaries of an in vitro heart, human or animal. By this hypothesis, different types of blood samples (for example, blood from a normal control compared with thalassemic blood) would give rise to differences in $R_2$* in both the septum and those regions susceptible to artefacts from the GCV.

## *8.3 General Conclusion*

Magnetic resonance imaging is an important clinical tool for the non-invasive investigation of soft tissue in the body. While MRI has been used as a qualitative and visual tool for assessment and diagnosis of many conditions, the technology has great potential in the quantification of relaxation properties in tissue. The clinical utility of relaxation rate quantification has been demonstrated in the quantification of body iron stores, owing to the relationship between proton relaxation rates and the presence of paramagnetic iron [7]. The digital nature of MR images provides an additional dimension of information, as the information contained in each voxel is not only intensity but also a spatial location. The effective representation of this information could potentially provide many useful clinical tools.

The work in this thesis has demonstrated that the mapping of relaxation rates in tissue provides insight into not only the mean relaxation properties of tissue but also the relationships between spatial variations of these properties and tissue properties. The development of an accurate method of measuring relaxation rates is essential to obtaining an accurate representation of the spatial information contained in the MR images. Potential clinical applications in the assessment of liver fibrosis and the quantification of artefacts in relaxometry analyses have been presented. Without the use

of voxel-by-voxel analyses of the MR images, the spatial information that is central to these two applications would be lost. The utility of spatial distributions of relaxation rates could extend well beyond the two applications discussed in this work to any other pathology where the texture or internal physical properties of the tissue are expected to change.

# References


1. Bloch, F., *Nuclear induction.* Physical Review, 1946. **70**(7): p. 460-474.
2. Purcell, E.M., H.C. Torrey, and R.V. Pound, *Resonance absorption by nuclear magnetic moments in a solid.* Physical Review, 1946. **69**(37).
3. Damadian, R., *Tumor detection by nuclear magnetic resonance.* Science, 1971. **171**: p. 1151-1153.
4. Lauterbur, P.C., *Image formation by induced local interactions: Examples employing numclear magnetic resonance.* Nature, 1973. **242**(5394): p. 190-191.
5. Kumar, A., D. Welti, and R.R. Ernst, *NMR Fourier Zeumatography.* Journal of Magnetic Resonance, 1975. **18**(1): p. 69-83.
6. Mansfield, P. and A.A. Maudsley, *Medical imaging by NMR.* British Journal of Radiology, 1977. **50**(591): p. 188-194.
7. Engelhardt, R., et al., *Liver iron quantification: studies in aqueous iron solutions, iron overloaded rats and patients with hereditary hemochromatosis.* Magnetic Resonance Imaging, 1994. **12**(7): p. 999-1007.
8. St Pierre, T.G., P.R. Clark, and W. Chua-anusorn, *Single spin-echo proton transverse relaxometry of iron-loaded liver.* NMR In Biomedicine, 2004. **17**(7): p. 446-458.
9. Aisen, P., C. Enns, and M. Wessling-Resnick, *Chemistry and biology of eukaryotic iron metabolism.* International Journal of Biochemistry & Cell Biology., 2001. **33**(10): p. 940-59.
10. Hoffbrand, A.V., et al., *Effect of iron deficiency and desferrioxamine on DNA synthesis in human cells.* British Journal of Haematology, 1976. **33**: p. 517-26.
11. Gutteridge, J.M.C. and B. Halliwell, *Iron toxicity and oxygen radicals*, in *Clinical Haematology*, C. Hershko, Editor. 1989, Bailliere Tindall: London. p. 195-255.
12. Bothwell, T.H., *Overview and mechanisms of iron regulation.* Nutrition Reviews, 1995. **53**(9): p. 237-245.
13. Andrews, N.C., *Disorders of iron metabolism.* New England Journal of Medicine, 1999. **341**(26): p. 1986-1995.
14. Feder, J., et al., *A novel MHC class I-like gene is mutated in patients with hereditary hemochromatosis.* Nature Genetics, 1996. **13**: p. 399-408.
15. Pippard, M.J., *Secondary Iron Overload*, in *Iron metabolism in health and disease*, J.H. Brock, et al., Editors. 1994, W.B. Saunders Company Ltd.: London. p. 271.
16. Fucharoen, S. and P. Winichagoon, *Thalassemia in SouthEast Asia: problems and strategy for prevention and control.* Southeast Asian Journal of Tropical Medicine & Public Health., 1992. **23**(4): p. 647-55.
17. Papanikolaou, G. and K. Pantopoulos, *Iron metabolism and toxicity.* Toxicology and Applied Pharmacology, 2005. **202**(2): p. 199-211.
18. Olivieri, N.F. and G.M. Brittenham, *Iron-chelating therapy and the treatment of thalassemia.* Blood, 1997. **89**(3): p. 739-61.
19. St. Pierre, T.G., et al., *A New Model for Predicting Venesection Therapy Requirements in Hereditary Hemochromatosis Using Non-Invasive Liver Iron Concentration Measurement.* ASH Annual Meeting Abstracts, 2005. **106**(11): p. 3596-.
20. Kreeftenberg, H.G., et al., *Measurement of iron in liver biopsies - a comparison of three analytical methods.* Clinica Chimica Acta., 1984. **144**(2-3): p. 255-62.
21. Emond, M.J., et al., *Quantitative study of the variability of hepatic iron concentrations.* Clinical Chemistry, 1999. **45**(3): p. 340-346.

22. Butensky, E., et al., *Variability in hepatic iron concentration in percutaneous needle biopsy specimens from patients with transfusional hemosiderosis.* American Journal Of Clinical Pathology, 2005. **123**(1): p. 146-152.
23. Regev, A., et al., *Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection.* American Journal of Gastroenterology., 2002. **97**(10): p. 2614-8.
24. Afdhal, N.H. and D. Nunes, *Evaluation of liver fibrosis: A concise review.* American Journal of Gastroenterology, 2004. **99**(6): p. 1160-1174.
25. Nord, J.H., *Biopsy diagnosis of cirrhosis: blind percutaneous versus guided direct vision technique. A review.* Gastrointestinal Endoscopy, 1982. **28**(102-104).
26. Perrault, J., et al., *Liver biopsy: complications in 1000 inpatients and outpatients.* Gastroenterology., 1978. **74**(1): p. 103-6.
27. Janes, C.H. and K.D. Lindor, *Outcome of patients hospitalized for complications after outpatient liver biopsy.* Annals of Internal Medicine, 1993. **118**: p. 96-98.
28. Ishak, K., et al., *Histological grading and staging of chronic hepatitis.* Journal of Hepatology., 1995. **22**(6): p. 696-9.
29. Bedossa, P., D. Dargere, and V. Paradis, *Sampling variability of liver fibrosis in chronic hepatitis C.* Hepatology, 2003. **38**(6): p. 1449-1457.
30. Goldin, R.D., et al., *Intra-observer and inter-observer variation in the histopathological assessment of chronic viral hepatitis.* Journal of Hepatology., 1996. **25**(5): p. 649-54.
31. Pinzani, M. and K. Rombouts, *Liver fibrosis: from the bench to clinical targets.* Digestive and Liver Disease, 2004. **36**(4): p. 231-242.
32. Poynard, T., et al., *Fibrosis in patients with chronic hepatitis C: Detection and significance.* Seminars in Liver Disease, 2000. **20**(1): p. 47-55.
33. Poynard, T., et al., *Impact of interferon alfa-2b and ribavirin on progression of liver fibrosis in patients with chronic hepatitis C.* Hepatology, 2000. **32**(5): p. 1131-1137.
34. Dufour, J.F., R. DeLellis, and M.M. Kaplan, *Regression of hepatic fibrosis in hepatitis C with long-term interferon treatment.* Digestive Diseases And Sciences, 1998. **43**(12): p. 2573-2576.
35. Friedman, S.L., *The cellular basis of hepatic fibrosis - mechanisms and treatment strategies.* The New England Journal of Medicine, 1993. **328**(25): p. 1828-1835.
36. Pilette, C., et al., *Histopathological evaluation of liver fibrosis: quantitative image analysis vs semi-quantitative scores. Comparison with serum markers.* Journal of Hepatology., 1998. **28**(3): p. 439-46.
37. Feldmann, G., *Critical analysis of the methods used to morphologically quantify hepatic fibrosis.* Journal of Hepatology, 1995. **22**((Suppl. 2)): p. 49-54.
38. Zaitoun, A.M., et al., *Quantitative assessment of fibrosis and steatosis in liver biopsies from patients with chronic hepatitis C.* Journal of Clinical Pathology., 2001. **54**(6): p. 461-5.
39. Sakaida, I., et al., *Quantitative analysis of liver fibrosis and stellate cell changes in patients with chronic hepatitis C after interferon therapy.* American Journal of Gastroenterology., 1999. **94**(2): p. 489-96.
40. Masseroli, M., et al., *Automatic quantification of liver fibrosis: design and validation of a new image analysis method: comparison with semi-quantitative indexes of fibrosis.* Journal of Hepatology, 2000. **32**: p. 453-464.
41. Dahab, G.M., et al., *Digital quantification of fibrosis in liver biopsy sections: Description of a new method by Photoshop software.* Journal Of Gastroenterology And Hepatology, 2004. **19**(1): p. 78-85.

42. Imbert-Bismut, F., et al., *Biochemical markers of liver fibrosis in patients with hepatitis C virus infection.* The Lancet, 2001. **357**: p. 1069-1075.
43. Naveau, S., et al., *Biomarkers for the prediction of liver fibrosisin patients with chronic alcoholic liver disease.* Clin Gastroenterol Hepatol, 2005. **3**: p. 167-174.
44. Imbert-Bismut, F., et al., *Intra-laboratory analytical variability of biochemical markers of fibrosis (Fibrotest) and activity (Actitest) and reference ranges in healthy blood donors.* Clinical Chemistry And Laboratory Medicine, 2004. **42**(3): p. 323-333.
45. Ngo, Y., et al., *A prospective analysis of FibroTest-ActiTest-FibroSure prognostic value in patients with chronic hepatitis C (CHC).* Hepatology, 2005. **42**: p. 438A.
46. Rossi, E., et al., *Validation of the FibroTest biochemical markers score in assessing liver fibrosis in hepatitis C patients.* Clinical Chemistry., 2003. **49**(3): p. 450-454.
47. Myers, R.P., et al., *Biochemical markers of liver fibrosis: a comparison with historical features in patients with chronic hepatitis C.* American Journal of Gastroenterology., 2002. **97**(9): p. 2419-25.
48. Castera, L., et al., *Prospective comparison of transient elastography, fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C.* Gastroenterology, 2005. **128**(2): p. 343-350.
49. Adams, L.A., et al., *Hepascore: An accurate validated predictor of liver fibrosis in chronic hepatitis C infection.* Clinical Chemistry, 2005. **51**(10): p. 1867-1873.
50. Forns, X., et al., *Identification of chronic hepatitis C patients without hepatic fibrosis by a simple predictive model.* Hepatology, 2002. **36**(4): p. 986-992.
51. Wai, C.T., et al., *A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C.* Hepatology, 2003. **38**(2): p. 518-526.
52. Leroy, V., et al., *Circulating matrix metalloproteinases 1, 2, 9 and their inhibitors TIMP-1 and TIMP-2 as serum markers of liver fibrosis in patients with chronic hepatitis C: Comparison with PIIINP and hyaluronic acid.* American Journal Of Gastroenterology, 2004. **99**(2): p. 271-279.
53. Boeker, K.H.W., et al., *Diagnostic potential of circulating TIMP-1 and MMP-2 as markers of liver fibrosis in patients with chronic hepatitis C.* Clinica Chimica Acta, 2002. **316**(1-2): p. 71-81.
54. Callewaert, N., et al., *Noninvasive diagnosis of liver cirrhosis using DNA sequencer-based total serum protein glycomics.* Nature Medicine, 2004. **10**(4): p. 429-434.
55. Lichtinghagen, R. and M.J. Bahr, *Noninvasive diagnosis of fibrosis in chronic liver disease.* Expert Review of Molecular Diagnostics, 2004. **4**(5): p. 715-726.
56. George, J., *Biochemical markers of hepatic fibrogenesis: single measurements are not reliable enough to replace liver biopsy - Editorial.* Journal of Gastroenterology and Hepatology, 2000. **15**: p. 819-821.
57. Yeh, W.C., et al., *Elastic modulus measurements of human liver and correlation with pathology.* Ultrasound In Medicine And Biology, 2002. **28**(4): p. 467-474.
58. Sandrin, L., et al., *Transient elastography: A new noninvasive method for assessment of hepatic fibrosis.* Ultrasound in Medicine and Biology, 2003. **29**(12): p. 1705-1713.
59. Ziol, M., et al., *Noninvasive assessment of liver fibrosis by measurement of stiffness in patients with chronic hepatitis C.* Hepatology, 2005. **41**(1): p. 48-54.
60. Saito, H., et al., *Efficacy of non-invasive elastometry on staging of hepatic fibrosis.* Hepatology Research, 2004. **29**(2): p. 97-103.

61. Kreft, B., et al., *Evaluation of different models of experimentally induced liver cirrhosis for MRI research with correlation to histopathologic findings.* Investigative Radiology., 1999. **34**(5): p. 360-366.
62. Thomsen, C., et al., *Prolonged T1 in patients with liver cirrhosis: an in vivo MRI study.* Magnetic Resonance Imaging, 1990. **8**(5): p. 599-604.
63. Aguirre, D.A., et al., *Accurate non-invasive diagnosis of liver fibrosis using MR imaging.* Hepatology, 2004. **40**: p. 707A.
64. Aube, C., et al., *Diagnosis and quantification of hepatic fibrosis with diffusion weighted MR imaging: preliminary results.* Journal De Radiologie, 2004. **85**(3): p. 301-306.
65. Blanc, J.F., et al., *Investigation of liver fibrosis in clinical practice.* Hepatology Research, 2005. **32**(1): p. 1-8.
66. Haktanir, A., et al., *Value of Doppler sonography in assessing the progression of chronic viral hepatitis and in the diagnosis and grading of cirrhosis.* Journal of Ultrasound in Medicine, 2005. **24**(3): p. 311-321.
67. Aube, C., et al., *New Doppler ultrasound signs improve the noninvasive diagnosis of cirrhosis or severe liver fibrosis.* European Journal of Gastroenterology & Hepatology, 2004. **16**(8): p. 743-751.
68. Hirata, M., et al., *Noninvasive diagnosis of the degree of hepatic fibrosis using ultrasonography in patients with chronic liver disease due to hepatitis C virus.* European Journal of Clinical Investigation, 2001. **31**(6): p. 528-535.
69. Barbaro, B., et al., *Doppler sonographic assessment of functional response of the right and left portal venous branches to a meal.* Journal of Clinical Ultrasound, 1999. **27**(2): p. 75-80.
70. Brittenham, G.M., *Noninvasive methods for the early detection of hereditary hemochromatosis.* Annals of the New York Academy of Sciences., 1988. **526**: p. 199-208.
71. Bauman, J.H. and J.W. Harris, *Estimation of hepatic iron stores by vivo measurement of magnetic susceptibility.* Journal of Laboratory & Clinical Medicine., 1967. **70**(2): p. 246-57.
72. Brittenham, G.M., et al., *Magnetic-susceptibility measurement of human iron stores.* New England Journal of Medicine., 1982. **307**(27): p. 1671-5.
73. Farrell, D., et al., *Magnetic measurement of human iron stores.* IEEE Transactions on Magnetics, 1980. **MAG-16**(5): p. 818-823.
74. Carneiro, A.A.O., et al., *An alternating current superconductor susceptometric system to evaluate liver iron overload.* Review Of Scientific Instruments, 2003. **74**(6): p. 3098-3103.
75. Fischer, R., et al., *Liver iron quantification in the diagnosis and therapy control of iron overload patients,* in *Biomagnetism: clinical aspects,* M. Hoke, et al., Editors. 1992, Elsevier Science Publishers B.V.: Amsterdam. p. 585-588.
76. Fischer, R., et al., *Assessment of iron staves in children with transfusion siderosis by biomagnetic liver susceptometry.* American Journal of Hematology, 1999. **60**(4): p. 289-299.
77. Nielsen, P., et al., *Non-invasive liver iron quantification by SQUID-biosusceptometry and serum ferritin iron as new diagnostic parameters in hereditary hemochromatosis.* Blood Cells Molecules and Diseases, 2002. **29**(3): p. 451-458.
78. Nielsen, P., et al., *Serum ferritin iron in iron overload and liver damage: Correlation to body iron stores and diagnostic relevance.* Journal of Laboratory and Clinical Medicine, 2000. **135**(5): p. 413-418.

79. Nielsen, P., et al., *Liver iron stores in patients with secondary haemosiderosis under iron chelation therapy with deferoxamine or deferiprone.* British Journal of Haematology., 1995. **91**(4): p. 827-833.
80. 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.* British Journal of Haematology, 2003. **121**(6): p. 938-948.
81. Moreira, M., L.O. Murta, and O. Baffa, *Imaging ferromagnetic tracers with an ac biosusceptometer.* Review Of Scientific Instruments, 2000. **71**(6): p. 2532-2538.
82. Cora, L.A., et al., *AC biosusceptometry in the study of drug delivery.* Advanced Drug Delivery Reviews, 2005. **57**(8): p. 1223-1241.
83. Piga, A., et al., *Comparison of LIC Obtained from Biopsy, BLS and R2-MRI in Iron Overloaded Patients with {beta}-Thalassemia, Treated with Deferasirox (Exjade(R), ICL670).* ASH Annual Meeting Abstracts, 2005. **106**(11): p. 2689-.
84. Carneiro, A.A.O., et al., *Liver iron concentration evaluated by two magnetic methods: Magnetic resonance imaging and magnetic susceptometry.* Magnetic Resonance In Medicine, 2005. **54**(1): p. 122-128.
85. Liang, Z. and P.C. Lauterbur, *Principles of magnetic resonance imaging: a signal processing perspective*. 2000, New York: IEEE Press.
86. Haacke, E.M., et al., *Magnetic resonance imaging: physical principles and sequence design*. 1999, New York: Wiley.
87. Hahn, E.L., *Spin Echoes.* Physical Review, 1950. **80**(4): p. 580-594.
88. Vymazal, J., et al., *Iron uptake by ferritin: NMR relaxometry studies at low iron loads.* Journal Of Inorganic Biochemistry, 1998. **71**(3-4): p. 153-157.
89. Koenig, S.H. and R.D. Brown (III), *Relaxation of solvent protons by paramagnetic ions and its dependence on magnetic field and chemical environment: implications for NMR imaging.* Magnetic Resonance in Medicine., 1984. **1**(4): p. 478-95.
90. Gandon, Y., et al., *Hemochromatosis: diagnosis and quantification of liver iron with gradient-echo MR imaging.* Radiology, 1994. **193**(2): p. 533-8.
91. Bonkovsky, H.L., et al., *Usefulness and limitations of laboratory and hepatic imaging studies in iron-storage disease.* Gastroenterology, 1990. **99**(4): p. 1079-91.
92. Kaltwasser, J.P., et al., *Non-invasive quantification of liver iron-overload by magnetic resonance imaging.* British Journal of Haematology, 1990. **74**: p. 360-363.
93. Papakonstantinou, O.G., et al., *Assessment of liver iron overload by T2-quantitative magnetic resonance imaging: Correlation of T2-QMRI measurements with serum ferritin concentration and histologic grading of siderosis.* Magnetic Resonance Imaging, 1995. **13**(7): p. 967-977.
94. Gandon, Y., et al., *Non-invasive assessment of hepatic iron stores by MRI.* The Lancet, 2004. **363**: p. 357-362.
95. Anderson, L.J., et al., *Cardiovascular T2-star (T2*) magnetic resonance for the early diagnosis of myocardial iron overload.* European Heart Journal, 2001. **22**(23): p. 2171-2179.
96. Jensen, P.D., *Evaluation of iron overload.* British Journal of Haematology, 2004. **124**(6): p. 697-711.
97. St Pierre, T.G., et al., *Noninvasive measurement and imaging of liver iron concentrations using proton magnetic resonance.* Blood, 2005. **105**(2): p. 855-861.

98. Johnston, D.L., et al., *Assessment of tissue iron overload by nuclear magnetic resonance imaging.* American Journal of Medicine, 1989. **87**(1): p. 40-7.
99. Jensen, P.D., et al., *Evaluation of myocardial iron by magnetic resonance imaging during iron chelation therapy with deferrioxamine: indication of close relation between myocardial iron content and chelatable iron pool.* Blood, 2003. **101**(11): p. 4632-4639.
100. Mavrogeni, S.I., et al., *T2 relaxation time study of iron overload in b-thalassemia.* Magnetic Resonance Materials in Physics Biology and Medicine, 1998. **6**: p. 7-12.
101. Mavrogeni, S.I., et al., *A comparison of magnetic resonance imaging and cardiac biopsy in the evaluation of heart iron overload in patients with beta-thalassemia major.* European Journal Of Haematology, 2005. **75**(3): p. 241-247.
102. Borgna-Pignatti, C., et al., *Survival and disease complications in thalassemia majpr.* Annals of the New York Academy of Sciences, 1998. **850**: p. 227-231.
103. Buja, L.M. and W.C. Roberts, *Iron in the heart - Etiology and clinical significance.* The American Journal of Medicine, 1971. **51**: p. 209-221.
104. Henry, W.L., et al., *Echocardiographic abnormalities in patients with transfusion-dependent anemia and secondary myocardial iron deposition.* American Journal of Medicine, 1978. **1978**(64): p. 4.
105. Ehlers, K.H., et al., *Longitudinal study of cardiac function in thalassemia major.* Annals of the New York Academy of Sciences, 1980. **344**: p. 397-404.
106. Davis, B.A. and J.B. Porter, *Long-term outcome of continuous 24-hour deferoxamine infusion via indwelling intravenous catheters in high-risk beta-thalassemia.* Blood, 2000. **95**(4): p. 1229-1236.
107. Lombardo, T., et al., *Reversibility of the cardiac iron load in thalassemia patients: a clinical pathology study.* Haematologica, 2003. **6**(3): p. 336-343.
108. Salonen, J.T., et al., *High stored iron levels are associated with excess risk of myocardial infarction in eastern Finnish men.* Circulation, 1992. **86**: p. 803-811.
109. Porter, J.B., *Monitoring and treatment of iron overload: State of the art and new approaches.* Seminars In Hematology, 2005. **42**(2): p. S14-S18.
110. Gabutti, V. and A. Piga, *Results of long-term iron-chelation therapy.* Acta Haematologica, 1996. **95**: p. 26-36.
111. Anderson, L.J., et al., *Comparison of effects of oral deferiprone and subcutaneous desferrioxamine on myocardial iron concentrations and ventricular function in beta-thalassaemia.* Lancet, 2002. **360**(9332): p. 516-520.
112. Olivieri, N.F., et al., *Long-term safety and effectiveness of iron-chelation therapy with deferiprone for thalassemia major.* New England Journal of Medicine, 1998. **339**(7): p. 417-423.
113. Hoffbrand, A.V., et al., *Long-term trial of deferiprone in 51 transfusion-dependent iron overloaded patients.* Blood., 1998. **91**(1): p. 295-300.
114. Pennell, D.J., et al., *Randomized controlled trial of deferiprone or deferoxamine in beta-thalassemia major patients with asymptomatic myocardial siderosis.* Blood, 2006. **107**(9): p. 3738-3744.
115. Borgna-Pignatti, C., et al., *Cardiac morbidity and mortality in deferoxamine- or deferiprone-treated patients with thalassemia major.* Blood, 2006. **107**(9): p. 3733-3737.
116. Olson, L.J., et al., *Cardiac iron depositin in idiopathic hemochromatosis: histologic and analytic assessment of 14 hearts from autopsy.* Journal of the American College of Cardiology, 1987. **10**: p. 1239-1243.
117. Liu, P., et al., *Quantification of cardiac and tissue iron by nuclear magnetic resonance relaxometry in a novel murine thalassemia-cardiac iron overload model.* Canadian Journal of Cardiology, 1996. **12**(2): p. 155-164.

118. Wood, J.C., et al., *Cardiac iron determines cardiac T2*, T2, and T1 in the gerbil model of iron cardiomyopathy.* Circulation, 2005. **112**(4): p. 535-543.
119. Wang, Z.Y.J., et al., *1/T2 and magnetic susceptibility measurements in a gerbil cardiac iron overload model.* Radiology, 2005. **234**(3): p. 749-755.
120. Jensen, P.D., et al., *Indirect evidence for the potential ability of magnetic resonance imaging to evaluate the myocardial iron content in patients with transfusional iron overload.* Magnetic Resonance Materials in Physics Biology and Medicine, 2001. **12**(2-3): p. 153-166.
121. Jensen, P.D., et al., *Non-invasive assessment of tissue iron overload in the liver by magnetic resonance imaging.* British Journal of Haematology, 1994. **87**(1): p. 171-84.
122. Mavrogeni, S.I., et al., *Myocardial iron deposition in beta-Thalassemia studied by magnetic resonance imaging.* International Journal of Cardiac Imaging, 1998. **14**(2): p. 117-122.
123. Papanikolaou, N., et al., *Non-invasive myocardial iron assessment in thalassaemic patients - T2 relaxometry and magnetization transfer ratio measurements.* Acta Radiologica, 2000. **41**(4): p. 348-351.
124. Voskaridou, E., et al., *Magnetic resonance imaging in the evaluation of iron overload in patients with beta thalassaemia and sickle cell disease.* British Journal of Haematology, 2004. **126**: p. 736-742.
125. Wood, J.C., et al., *Myocardial iron loading in transfusion-dependent thalassemia and sickle cell disease.* Blood, 2004. **103**(5): p. 1934-1936.
126. Westwood, M., et al., *A single breath-hold multiecho T2* cardiovascular magnetic resonance technique for diagnosis of myocardial iron overload.* Journal of Magnetic Resonance Imaging, 2003. **18**(1): p. 33-39.
127. Westwood, M., et al., *Interscanner reproducibility of cardiovascular magnetic resonance T2* measurements of tissue iron in thalassemia.* Journal of Magnetic Resonance Imaging, 2003. **18**: p. 616-620.
128. Galia, M., et al., *Potential myocardial iron content evaluation by magnetic resonance imaging in thalassemia major patients treated with deferoxamine or deferiprone during a randomized multicenter prospective clinical study.* Hemoglobin, 2003. **27**(2): p. 63-76.
129. Anderson, L.J., et al., *Myocardial iron clearance during reversal of siderotic cardiomyopathy with intravenous desferrioxamine: a prospective study using T2*cardiovascular magnetic resonance.* British Journal of Haematology, 2004. **127**(3): p. 348-355.
130. Reeder, S.B., *In vivo measurement of T2* and field inhomogeneity maps in the human heart at 1.5T.* Magnetic Resonance in Medicine, 1998. **39**: p. 988-998.
131. Bauer, W.R., et al., *Theory of the BOLD effect in the capillary region: an analytical approach for the determination of T2* in the capillary network of the myocardium.* Magnetic Resonance in Medicine, 1999. **41**: p. 51-62.
132. Koelher, S., et al., *Visualization of myocardial microstructure using high resolution T2* imaging at high magnetic field.* Magnetic Resonance in Medicine, 2003. **49**: p. 371-375.
133. Wacker, C., et al., *Changes in myocardial oxygenation and perfusion under pharmacological stress with dipyridamole: assessment using T2* and T1 measurements.* Magnetic Resonance in Medicine, 1999. **41**: p. 686-695.
134. Zhang, J.B. and G.A. Krinsky, *Iron-containing nodules of cirrhosis.* Nmr In Biomedicine, 2004. **17**(7): p. 459-464.
135. Freeborough, P.A. and N.C. Fox, *MR image texture analysis applied to the diagnosis and tracking of Alzheimer's disease.* IEEE Transactions on Medical Imaging, 1998. **17**(3): p. 475-479.

136. Pavlopoulos, S., et al., *Fuzzy neural network-based texture analysis of ultrasonic images.* IEEE Engineering in Medicine and Biology, 2000. **19**: p. 39-47.
137. Haralick, R.M., K. Shanmugam, and I.h. Dinstein, *Textural features for image classification.* IEEE Transactions on Systems, Man and Cybernetics, 1973. **3**: p. 610-621.
138. Clark, P.R. and T.G.S. Pierre, *Quantitative mapping of transverse relaxivity (1/T2) in hepatic iron overload: a single spin-echo imaging methodology.* Magnetic Resonance Imaging, 2000. **18**: p. 431-438.
139. Clark, P.R., W. Chua-anusorn, and T.G. St Pierre, *Reduction of respiratory motion artifacts in transverse relaxation rate (R2) images of the liver.* 2004. **28**: p. 69-76.
140. Hendrick, R.E., *Image conrast and noise*, in *Magnetic Resonance Imaging*, D.D. Stark and W.G. Bradley, Editors. 1988, The C. V. Mosby Company: St. Louis. p. 67-72.
141. Cole, W., A. LeBlanc, and S. Jhingran, *The origin of biexponential $T_2$ relaxation in muscle water.* Magnetic Resonance Imaging, 1993. **29**(1): p. 19-24.
142. 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.* Magnetic Resonance Imaging, 2003. **21**: p. 519-530.
143. Castellano, G., et al., *Texture analysis of medical images.* Clinical Radiology, 2004. **59**(12): p. 1061-1069.
144. Scott, D.W., *On optimal and data-based histograms.* Biometrika, 1979. **66**: p. 605-610.
145. Pietrangelo, A., et al., *Enhanced hepatic collagen type I mRNA expression into fat-storing cells in a rodent model of hemochromatosis.* Hepatology, 1994. **19**(3): p. 714-721.
146. Clark, P.R., W. Chua-anusorn, and T.G. St Pierre, *Proton transverse relaxation rate (R-2) images of iron-loaded liver tissue: mapping local tissue iron concentrations with MRI.* Magnetic Resonance In Medicine, 2003. **49**(3): p. 572-575.
147. Bolondi, L., et al., *Characterization of small nodules in cirrhosis by assessment of vascularity: The problem of hypovascular hepatocellular carcinoma.* Hepatology, 2005. **42**(1): p. 27-34.
148. Olivieri, N.F., *The beta-thalassemias.* New England Journal of Medicine., 1999. **341**(2): p. 99-109.
149. Hosmer, D.W. and S. Lemeshow, *Applied Logistic Regression*. Second ed. Wiley Series in Probability and Statistics. 2000: Wiley-Interscience.
150. Peduzzi, P., et al., *A simulation study of the number of events per variable in logistic regression analysis.* Journal of Clinical Epidemiology, 1996. **49**(12): p. 1373-1379.
151. Freedman, L.S. and D. Pee, *Return to a note on screening regression equations.* The American Statistician, 1989. **43**(4): p. 279-282.
152. Concato, J., et al., *Importance of events per independent variable in proportional hazards analysis I. Background, goals and general strategy.* Journal of Clinical Epidemiology, 1995. **48**(12): p. 1495-1501.
153. Peduzzi, P., et al., *Importance of events per independent variable in proportional hazards regression analysis: II. Accuracy and precision of regression estimates.* Journal of Clinical Epidemiology, 1995. **48**(12): p. 1503-1510.

154. Harrell, F.E., et al., *Regression models for prognostic prediction: advantages, problems and suggested solutions.* Cancer Treatment Reports, 1985. **69**: p. 1071-1077.
155. Moser, E., E. Winklmayr, and M. Krssak, *On the correlation between tissue hydration state and proton NMR relaxation rates in experimental liver transplantation.* NMR in Biomedicine, 1997. **10**: p. 143-150.
156. Villeneuve, J.-P., et al., *Variability in hepatic iron concentration measurement from needle-biopsy specimens.* Journal of Hepatology, 1996. **25**(2): p. 172-177.
157. Bassett, M.L., J.W. Halliday, and P. Lawrie W, *Value of hepatic iron measurements in early hemochromatosis and determination of the critical iron level associated with fibrosis.* Hepatology, 1986. **6**(1): p. 24-29.
158. Moser, E., P. Holzmueller, and G. Gomiscek, *Liver tissue characterization by in vitro NMR: tissue handling and biological variation.* Magnetic Resonance in Medicine., 1992. **24**(2): p. 213-20.
159. Haacke, E.M., et al., *Imaging iron stores in the brain using magnetic resonance imaging.* Magnetic Resonance Imaging, 2005. **23**(1): p. 1-25.
160. Dinh, A. and A. Tyler, *Construction of gelatine phantoms for use in MR relaxometry (Personal Communication)*. 2005.
161. Farzaneh, F., S.J. Riederer, and N.J. Pelc, *Analysis of T2 limitations and off-resonance effects on spatial resolution and artifacts in echo planar imaging.* Magnetic Resonance in Medicine, 1990. **14**: p. 123-139.
162. Truong, T.-K., et al., *Three-dimensional numerical simulations of susceptibility-induced magnetic fiels inhomogeneities in the human head.* Magnetic Resonance Imaging, 2002. **20**: p. 759-770.
163. Olivieri, N.F., et al., *Survival In Medically Treated Patients With Homozygous beta-Thalassemia.* New England Journal of Medicine, 1994. **331**(9): p. 574-578.
164. Thomsen, C., et al., *Identification of patients with hereditary haemochromatosis by magnetic resonance imaging and spectroscopic relaxation time measurements.* Magnetic Resonance Imaging, 1992. **10**(6): p. 867-879.
165. Rocchi, E., et al., *Magnetic Resonance Imaging and Different Levels of Iron Overload in Chronic Liver Disease.* Hepatology, 1993. **17**(6): p. 997-1002.
166. Ghugre, N.R., et al., *Improved R2* measurements in myocardial iron overload.* Journal Of Magnetic Resonance Imaging, 2006. **23**(1): p. 9-16.
167. Westwood, M., et al., *A single breath-hold multiecho T2* cardiovascular magnetic resonance technique for diagnosis of myocardial iron overload.* Journal of Magnetic Resonance Imaging, 2003. **18**: p. 33-39.
168. Ho, C.H., *The effect of transfusion on cardiac function in patients with chronic anemia.* Transfusion, 1997. **37**(10): p. 1066-1069.
169. Ho, C.H., *The changes of cardiac ejection fraction up to 72 h after transfusion in patients with chronic anaemia.* Transfusion Medicine, 2001. **11**(3): p. 189-192.
170. Pepe, A., et al., *Evaluation of the efficacy of oral deferiprone in beta-thalassemia major by multislice multiecho T2*.* European Journal Of Haematology, 2006. **76**(3): p. 183-192.
171. Bauminger, E.R., et al., *Quantitative studies of ferritinlike iron in erythrocytes of thalassemia, sickle-cell anemia, and hemoglobin Hammersmith with Mossbauer spectroscopy.* Proceedings of the National Academy of Sciences of the United States of America, 1979. **76**(2): p. 939-43.
172. Yen, C.S., et al., *Mossbauer studies of Fe2+ in aggregated and anhydrated sickle cell hemoglobin.* Biochimica et Biophysica Acta, 1976. **453**(1): p. 231-239.

173. Moal, F., et al., *Fractal dimension can distinguish models and pharmacological changes in liver fibrosis in rats.* Hepatology, 2002. **36**: p. 840-849.

# Appendix – Texture Measures

A vast number of algorithms can potentially be used to quantify texture from grey-tone spatial dependence (GTSD) matrices. A set of fifteen texture measures was used in this work.

The normalised GTSD matrix $p(i,j)$ is used and is given by

$$p(i,j) = P(i,j)/R$$

where $P(i,j)$ is the GTSD matrix and $R$ is the number of pixel pairs in the image.

The following definitions are used in the expressions for the texture measures

$$p_x(i) = \sum_{j=1}^{N} p(i,j)$$

$$p_y(j) = \sum_{i=1}^{N} p(i,j)$$

$$p_{x+y}(k) = \sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j), \qquad \text{where } k = i + j = 2, 3, \ldots, 2N.$$

$$p_{x-y}(k) = \sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j), \qquad \text{where } k = i - j = 0, 1, \ldots, N-1.$$

The vectors $p_x(i)$ and $p_y(j)$ have means $\mu_x$, $\mu_y$ and standard deviations $\sigma_x$, $\sigma_y$.

The following measures of entropy are also used in the texture measures

$$H = -\sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j)\log[p(i,j)]$$

$$H_{XY1} = -\sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j)\log[p_x(i)p_y(j)]$$

$$H_{XY2} = -\sum_{i=1}^{N}\sum_{j=1}^{N} p_x(i)p_y(j)\log[p_x(i)p_y(j)]$$

$$H_X = \sum_{i=1}^{N} p_x(i) \log[p_x(i)]$$

$$H_Y = \sum_{j=1}^{N} p_y(j) \log[p_y(j)]$$

$$H_{X+Y} = -\sum_{i=2}^{2N} p_{x+y}(i) \log[p_{x+y}(i)]$$

$$H_{X-Y} = -\sum_{i=0}^{N-1} p_{x-y}(i) \log[p_{x-y}(i)]$$

The texture measures used in this thesis are defined as follows –

| | |
|---|---|
| Texture measure $t_1$ | $t_1 = \sum_{i=1}^{N} \sum_{j=1}^{N} [p(i,j)]^2$ |
| Texture measure $t_2$ | $t_2 = \sum_{i=1}^{N} i^2 p_{x-y}(i)$ |
| Texture measure $t_3$ | $t_3 = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} (ij) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$ |
| Texture measure $t_4$ | $t_4 = \sum_{i=1}^{N} \sum_{j=1}^{N} (i-j)^2 p(i,j)$ |
| Texture measure $t_5$ | $t_5 = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{1+(i-j)^2} p(i,j)$ |
| Texture measure $t_6$ | $t_6 = \sum_{i=2}^{2N} i p_{x+y}(i)$ |
| Texture measure $t_7$ | $t_7 = \sum_{i=2}^{2N} (i - t_8)^2 p_{x+y}(i)$ |
| Texture measure $t_8$ | $t_8 = H_{X+Y}$ |
| Texture measure $t_9$ | $t_9 = H$ |
| Texture measure $t_{10}$ | $t_{10} = \sum_{i=0}^{N-1} (i - t_{11})^2 p_{x-y}(i)$ |
| Texture measure $t_{11}$ | $t_{11} = H_{X-Y}$ |

| | |
|---|---|
| Texture measure $t_{12}$ | $t_{12} = \frac{H - H_{XY1}}{\max\{H_X, H_Y\}}$ |
| Texture measure $t_{13}$ | $t_{13} = \sqrt{1 - e^{-2(H_{XY2} - H)}}$ |
| Texture measure $t_{14}$ | $t_{14} = \sum_{i=0}^{N-1} i p_{x-y}(i)$ |
| Texture measure $t_{15}$ | $t_{15} = \sum_{i=1}^{N} \sum_{j=1}^{N} (i-j)^2 p(i,j)$ |