Applying quantitative image analysis to liver cancer



Elisabeth Pickles

Department of Engineering Science and Linacre College

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Michaelmas Term 2022

Acknowledgements

I am very grateful to my supervisors (Michael Brady, Daniel Bulte and Michael Pavlides) for their support and guidance. I thank them for their willingness to meet me and review my work frequently, and for their helpful feedback. Their faith in me and my work helped me keep going through the highs and lows of the project.

I am also very grateful for all the support I received from colleagues at Perspectum Ltd. The number of people who helped me with various parts of my project mean that I cannot mention them all here, but to name a few: Ged Ridgway, Paul Aljabar, Matt Robson, Alison Telford, Shravan Kumar, Rui Teixeira, Alex Bagur and Jacky Law.

I am also indebted to the team at OCMR, especially Ferenc Mozes, as well as the radiographers.

I collaborated with many people during the project, all of which I would extend my thanks to. Being part of the DeLIVER Consortium, led by Ellie Barnes, has been an experience that I am very grateful for. Sue Francis and Eleanor Cox from University of Nottingham have played a key role in much of the work in this thesis. I really appreciate their willingness to answer my questions and support my work. Jonathan Dillman and Jean Tkach from Cincinnati Children's Hospital have also played important roles. I am grateful to their careful review of everything I've sent to them: their comments are always very useful.

Finally, I would like to thank all my family and friends for their support. They have had to put up with me through the highs and lows, and there have been many, especially due to Covid-19 impacting most of my DPhil. Thank you so much to you all.

Table of Contents

1. INTRODUCTION 1

1.1. MOTIVATION 1
1.2. OBJECTIVES 5
1.3. OUTLINE OF CHAPTERS 7

2. HEPATOCELLULAR CARCINOMA AND IRON SPARING 9

2.1. THE LIVER 9
2.2. CIRRHOSIS 12
2.3. HEPATOCARCINOGENESIS 13
2.4. IMAGING OF HCC 15
2.5. HCC STAGING AND TREATMENT 18
2.6. IRON AND HEPATOCELLULAR CARCINOMA 19
2.6.1. IRON IN THE LIVER 19
2.6.2. IRON IN THE LIVER PARENCHYMA AND HCC DEVELOPMENT 19
2.6.3. IRON IN HCC 20
2.7. SUMMARY 23

3. MRI PHYSICS AND DETECTION OF IRON USING MRI 24

3.1. INTRODUCTION 24
3.2. SIGNAL GENERATION 24
3.3. SPATIAL ENCODING 28
3.4. MRI CONTRASTS 31
3.5. ACCELERATION IN MRI 32
3.5.1. PARTIAL FOURIER 32
3.5.2. PARALLEL IMAGING 33
3.5.3. COMPRESSED SENSING 35
3.6. QUANTITATIVE MRI 37
3.7. IRON AND MRI 38
3.8. MRI AND DETECTION OF IRON SPARING 46
3.8.1. MRI FOR DETECTION OF IRON SPARING 46
3.8.2. PROPOSAL 49
3.8.3. CONSIDERATIONS WHEN USING R_2* MAPPING FOR IRON SPARING 49
3.8.4. REQUIREMENTS FOR DETECTION OF IRON SPARING 53
3.9. SUMMARY 56

4. IRON CONVERSION METHODS: AN EVALUATION 57

4.1. INTRODUCTION 57
4.2. METHODS 58
4.2.1. PATIENTS 58
4.2.2. MRI PROTOCOL AND POST PROCESSING 59
4.2.3. 1.5 T LIC CONVERSION 60
4.2.4. 3 T LIC CONVERSION 61
4.2.5. STATISTICAL ANALYSIS 64
4.2.6. RELEVANT THRESHOLDS FOR AGREEMENT 64
4.3. RESULTS 65
4.4. DISCUSSION 72
4.5. SUMMARY 75

5. INITIAL EVALUATION OF R_2* SEQUENCES: PHANTOM 76

5.1. INTRODUCTION 76
5.2. INITIAL EVALUATION USING A PHANTOM 77
5.2.1. INTRODUCTION 77
5.2.2. METHOD 77

5.2.3. RESULTS 81
5.2.4. DISCUSSION 88
5.3. EVALUATION OF COMPRESSED SENSING USING A PHANTOM 93
5.3.1. INTRODUCTION 93
5.3.2. METHOD 94
5.3.3. RESULTS 96
5.3.4. DISCUSSION 102
5.4. SUMMARY 106

6. INITIAL EVALUATION OF R_2* SEQUENCES: *IN VIVO* 107

6.1. INTRODUCTION 107
6.2. METHOD 108
6.2.1. PATIENTS 108
6.2.2. MRI ACQUISITION 108
6.2.3. ANALYSIS AND STATISTICS 109
6.2.4. ASSESSMENT AGAINST CLINICALLY IMPORTANT THRESHOLDS 112
6.2.5. STATISTICAL TESTS 113
6.3. RESULTS 113
6.3.1. ROI ANALYSIS 113
6.3.2. SCORING 114
6.3.3. WITHIN SLICE HETEROGENEITY: INTER-QUARTILE RANGE 114
6.3.4. BETWEEN SLICE HETEROGENEITY 114
6.3.1. ASSESSMENT OF FIT 115
6.4. DISCUSSION 121
6.5. SUMMARY 127

7. OPTIMISATION OF R_2* SEQUENCES FOR IRON SPARING DETECTION 128

7.1. INTRODUCTION 128
7.2. MATERIALS AND METHODS 129
7.3. RESULTS 133
7.3.1. PHILIPS OPTIMISATION: 3 T 133
7.3.2. SIEMENS OPTIMISATION: 1.5 T 141
7.3.3. SIEMENS OPTIMISATION: 3 T 144
7.3.4. FINAL PROTOCOL 145
7.4. DISCUSSION 150
7.5. SUMMARY 153

8. EVALUATION OF OPTIMISED R_2* SEQUENCES IN HEALTHY VOLUNTEERS 154

8.1. INTRODUCTION 154
8.2. EVALUATION OF SEQUENCES AT 1.5 T AND 3 T 154
8.2.1. INTRODUCTION 154
8.2.2. METHODS 155
8.2.3. RESULTS 161
8.2.4. DISCUSSION 175
8.3. COMPARISON ACROSS PHILIPS AND SIEMENS SCANNERS *IN VIVO* 177
8.3.1. INTRODUCTION 177
8.3.2. METHODS 178
8.3.3. RESULTS 179
8.3.4. DISCUSSION 184
8.4. SUMMARY 187

9. EVALUATION OF OPTIMISED R_2* SEQUENCES IN "TUMOURS" 188

9.1. INTRODUCTION 188
9.2. EVALUATION OF SEQUENCES IN A PHANTOM 188

9.2.1. INTRODUCTION 188
9.2.2. METHODS 189
9.2.3. RESULTS 195
9.2.4. DISCUSSION 205
9.3. PATIENTS WITH HEPATOCELLULAR CARCINOMA 207
9.3.1. INTRODUCTION 207
9.3.2. METHODS 208
9.3.3. RESULTS 209
9.3.4. DISCUSSION 211
9.4. SUMMARY 213

10. SUMMARY AND OUTLOOK 214

10.1. IMPROVEMENTS TO R_2* MAPS 214
10.2. PHANTOM DEVELOPMENT 216
10.3. CONVERSION TO IRON IN TUMOURS 216
10.4. ALTERNATIVE TO R_2* MAPPING FOR DETECTING IRON SPARING 217
10.5. COMBINING WITH OTHER MRI PARAMETERS 217
10.6. CONCLUSION 218

11. REFERENCES 219

APPENDIX A ADDITIONAL FIGURES FROM CHAPTER 4 225

APPENDIX B ADDITIONAL FIGURES FROM CHAPTER 6 228

APPENDIX C ADDITIONAL FIGURES FROM CHAPTER 8 231

APPENDIX D SUPPLEMENTARY MATERIAL TO CHAPTER 9 232

Abstract

Applying quantitative image analysis to liver cancer

Elisabeth Pickles
Department of Engineering Science and Linacre College
University of Oxford
A thesis submitted for the degree of Doctor of Philosophy
Michaelmas Term 2022

Survival rates are poor for hepatocellular Carcinoma (HCC) compared to many other cancers, and death rates are increasing in countries such as the United States and United Kingdom. To improve outcomes, tumours need to be detected at a much earlier stage. Magnetic resonance imaging (MRI) may provide a method to detect tumours earlier, potentially combined with other biomarkers.

The hypothesis underlying this thesis is that the ability to measure iron in the liver accurately (that is, with low error) could provide an early marker for HCC development, as HCC tumours are thought to typically contain less iron than the liver (iron sparing). However, previously iron in the liver could only be measured using biopsy. MRI provides a non-invasive tool to measure iron.

Current MRI techniques for measuring iron are not suitable for the application of detecting iron sparing as they are optimised to detect iron overload in diseases such as haemochromatosis and thalassemia. For iron sparing, low iron concentrations, and small, localised regions of differences in iron need to be detected.

In this thesis I highlight the various limitations with existing techniques and describe the optimisation process to develop methods to quantify iron sparing. I verified that these could potentially be used in iron sparing using healthy volunteer data, phantom data, and patients. These MRI techniques can now be used in clinical trials for assessing if iron sparing is indeed a feature of HCC and thus whether MRI could be used as a surveillance tool to detect iron sparing in patients at-risk of HCC.

1. Introduction

1.1. Motivation

The liver is the largest solid organ in the body and performs over 500 vital functions, including the removal of toxins from blood, bile production (needed for the absorption of fats), albumin production (a protein required to keep the correct pressure to keep fluid in blood vessels, as well as transporting substances around the body) and storage of vitamins and minerals[1]. The liver is unique in having two blood supplies: arterial and venous. Arterial blood is supplied through the hepatic artery and accounts for 25-30% of the blood supply of the liver, whereas venous blood is supplied through the portal vein and accounts for 70-75% of the blood supply[2].

Repeated damage to the liver, for example by viral hepatitis, excess alcohol consumption and non-alcoholic steatohepatitis can result in cirrhosis. Cirrhosis is an advanced stage of liver disease and is defined as "a diffuse process characterized by fibrosis and the conversion of normal liver architecture into structurally abnormal nodules"[3] by the World Health Organisation (1977).

Primary liver cancer originates in the liver, in contrast to metastatic liver cancer, which occurs when cancer spreads from another organ to the liver, for example from the breast or colon. Primary liver cancer typically develops in patients with aetiologies such as viral hepatitis and alcoholic liver disease, and on a background of a cirrhotic liver. Hepatocellular carcinoma (HCC) is the most common type of primary liver cancer[4]. Most HCCs (70 – 90%) occur on a background of liver cirrhosis[5]. Currently, it is not clear whether cirrhosis causes HCC or whether they develop largely independently[6]. Generally, HCC develops more slowly than cirrhosis, and so appears later.

Primary liver cancer is most prevalent in Eastern and South Eastern Asia and Central Africa primarily due to the prevalence of Hepatitis B in those countries[4]. In Eastern Asia the age standardised incidence is 24.3 cases per 100,000 population[7]. The disease is less prevalent in Europe and North America and South Central Asia, with age standardised incidence of 7 and 9 cases per 100,000 population in Central and Western Europe and 7.8 cases per 100,000 population in North America (high income)[7]. However incidence rates are increasing in many European countries and the United States[7], for example the age-adjusted incidence rate in the United States was 2.8 per 100,000 in 1975 compared to 8.2 in 2019[8]. This may be due to the prevalence of hepatitis C[7], but may also be driven by the increase in non-alcoholic fatty liver disease[4].

Primary liver cancer has very poor survival rates: in 2015, primary liver cancer was the 4th most common cause of cancer mortality, and was estimated to have caused 810 000 deaths Worldwide[7]. The number of deaths from liver cancer is increasing. For example, in the United States death rates from primary liver and bile duct cancer rose by approximately 1.3 % per year from 2010 to 2019[8]. The five year survival rate for primary liver and bile duct cancer in the United States is approximately 20.8%[8]. This is despite curative treatments being available: if HCC is detected at a very early stage (one lesion <2 cm) and the patient has good liver function then 5-year survival rates following resection are 80-90 %[9]. These rates are similar for radiofrequency ablation, another common treatment type[9]. To improve survival rates, the major unmet need is to detect such tumours early, when curable treatment is possible. To that end robust and reliable methods are required which allow for detection of tumours as early as possible.

Surveillance for HCC in at-risk populations on a six monthly basis is recommended by medical associations and societies throughout the world[9–11]. Typically, at-risk patients are those with cirrhosis and, in some countries, non-cirrhotic hepatitis B carriers[6].

However, despite the widespread surveillance of at-risk individuals, survival rates remain poor, indicating limitations with the current methods used in surveillance. One such limitation is the lack of a tool with adequate sensitivity and specificity for early detection of HCC. Another may be the rising incidence of HCC in patients with non-alcoholic steatohepatitis (NASH) without cirrhosis[12], as such patients are not included in surveillance programmes.

Currently available tools are not adequate for the early detection of HCC, although there has been some progress in this area. Serological tests have been investigated, and Alpha-fetoprotein (AFP) is a commonly used biomarker in HCC. However, for early stage tumours, only 10-20% of patients have abnormal AFP levels[9], and AFP fluctuations can be caused by viral hepatitis flares. Currently, societies such as European Association for the Study of the Liver (EASL) deem the evidence for using AFP for surveillance to be insufficient[9]. Ultrasound on a six monthly basis is the more common mode of surveillance in at-risk individuals, as proposed in the current EASL clinical guidelines[9]. However ultrasound only has 63% sensitivity[13] for early HCC. Other imaging methods with potential for early detection include computerised tomography (CT) and magnetic resonance imaging (MRI). Contrast-enhanced MRI (and/or CT) is already used for diagnosis of HCC[6 9 10]. The differences in blood flow between the liver parenchyma and tumour enable the identification of the tumour: during the arterial phase, when the blood flow is through the hepatic artery, the tumour will uptake more contrast relative to the liver, as it is primarily supplied by the hepatic artery, whereas during the portal venous phase, when the blood flow is through the portal vein, the tumour will show wash out and the liver tissue will have more contrast[6]. MRI shows greater soft tissue contrast and allows for the assessment of additional characteristics of tumours relative to CT. Although MRI

is the imaging of choice for diagnosis of HCC[14], CT may still be the chosen method for HCC imaging in less-specialized centres.

CT uses ionising radiation, so it is not suitable as a surveillance tool. Contrast-enhanced MRI is also not recommended for HCC surveillance due to the safety concern associated with gadolinium-based contrast agent retention and accumulation in human organs including the brain[15] following repeated administration of contrast in surveillance programmes[9]. Additionally there is a lack of evidence of the cost-effectiveness of using contrast-enhanced MRI for surveillance[10], although the results from one study comparing ultrasound with MRI with liver-specific contrast in patients found that MRI with liver-specific contrast may be more cost-effective than ultrasound[16].

I propose that a combination of markers, including MRI without contrast, could provide a tool for robust and reliable early detection of HCC in at-risk individuals. A feature of HCC which I hypothesise could be used for early detection is that the tumour may manifest iron sparing (lower levels of iron) compared to the liver parenchyma (liver tissue)[6,17–19]. However currently the number of studies investigating iron sparing in HCC are relatively few, with some conflicting results[17–21]. I hypothesise that a non-invasive tool, such as MRI, could be used to firstly verify if iron sparing does occur in the majority of tumours, and if so, that it can be used as a component tool for early detection of HCC.

Several methods using MRI to non-invasively measure liver iron have been developed[22], including a technique called R_2* mapping (explained in more detail in Chapter 3), but they are not in widespread use for detecting iron sparing in tumours and are currently not optimised for this application. Development of techniques to detect iron sparing is required. A challenge of MRI is that there are a number of different manufacturers who supply MRI machines, such as Philips and Siemens, and their hardware and software are slightly different. Additionally, MRI scanners can be of

different field strengths, typically measured in tesla (T). Both 1.5 T and 3 T scanners are in widespread clinical use. Therefore, any tool developed needs to work across scanners produced by different manufacturers and with different field strengths.

1.2. Objectives

The objective of this work was to investigate the proposition that quantitative MRI techniques could improve the early detection of HCC. The hypothesis underpinning this thesis in particular was that MRI sequences can be used to detect iron sparing. This has primarily involved developing and validating 3D R_2* mapping sequences.

This has necessitated:

1. Evaluation of 3 T R_2* to liver iron concentration (LIC) conversion techniques: it is important that 1.5 T R_2*-LIC and 3 T R_2*-LIC give the same results so they can be used alongside each other, both for use in this project focused on iron sparing, but also in wider applications such as monitoring iron in patients with thalassemia. 1.5 T R_2*-LIC is better validated, so it is important that 3 T R_2*-LIC matches 1.5 T R_2*-LIC. I compared techniques and gave recommendations on which conversion method should be used for 3 T R_2*. *See Chapter 4.*
2. Evaluation of existing/slightly modified 3D R_2* mapping sequences: these are becoming more widespread, however further work is required to validate that they give the same results as traditional 2D R_2* mapping sequences. Limitations were identified, particularly in the case of the novel applications of detecting iron sparing. Of importance, use of compressed sensing was evaluated, as there has been one study which suggested compressed sensing should not be used for R_2* mapping. *See Chapter 5 and Chapter 6.*

3. Optimisation of 3D R_2* mapping sequences for iron sparing detection: 3D R_2* sequences have primarily been used for detecting iron overload. To make them suitable for detection of iron sparing in HCC numerous modifications were required. This was performed on two scanner vendors (Siemens and Philips) and at two field strengths (1.5 T and 3 T) *See Chapter 7*
4. Validation of 3D R_2* mapping sequences for iron sparing detection in HCC: Healthy volunteers were scanned across both field strengths and vendors, to assess the sequence against the requirements for detecting iron sparing. A novel phantom was created to mimic HCCs to enable further validation against the requirements. Finally, a small cohort of patients with early HCC were scanned. *See Chapter 8 and Chapter 9.*

The major contribution of my work has been to show that it is possible to measure small, localised changes in iron using MRI, so that iron sparing associated with HCC can be detected. The sequences developed will be used in "The detection of Small Early Liver cancer with Natural history follow up" (SELINA) study, which will investigate techniques for early detection of liver cancer, and potentially the "Prospective Cohort for Early Detection of Liver Cancer" (PEARL) study. These studies are part of a CRUK funded program, DELIVER (Early Detection of Hepatocellular Liver Cancer).

I have made other contributions to the field including recommendations on 3 T-R_2* LIC conversion methods and assessment of, and recommendations for, 3D R_2* mapping sequences, in particular the assessment of compressed sensing. These contributions have widespread implications for clinicians and researchers, as R_2* mapping is becoming increasingly used for detecting iron overload in the clinic. Additionally, I have developed a novel phantom to mimic HCCs which could be replicated for other applications.

1.3. Outline of Chapters

Chapters 2 and 3 outline the background to this thesis. Chapter 2 provides the clinical and biological context, including the key background on iron sparing in HCC and the evidence in the literature for this. Chapter 3 covers the magnetic resonance (MR) physics context. This is followed by an explanation of how MRI can be used to detect iron, and the current literature on detecting iron sparing in MRI. It finishes with a summary and justification of requirements for a sequence that could enable detection of iron sparing.

Chapter 4 covers a study to investigate liver iron conversion techniques at 3 T. In this work participants were scanned at both 1.5 T and 3 T with sequences for R_2* mapping. I converted R_2* to iron using a variety of different methods and compared these between 1.5 T and 3 T, to assess which 3 T method gave the most similar result to 1.5 T (which is better validated). This informed me which 3 T-R_2* LIC conversion to use in future work in this project.

Chapter 5 and Chapter 6 describe evaluation of 3D R_2* mapping sequences in a phantom and patients with cirrhosis respectively. They outline the potential limitations of the sequences for detecting iron sparing. These chapters also give an evaluation of the use of compressed sensing in the phantom and *in vivo*, and an indication of whether it is acceptable to use compressed sensing in 3D R_2* mapping sequences.

Chapter 7 describes the optimisation of 3D R_2* mapping sequences to endeavour to reach the requirements for detecting iron sparing, as given in Chapter 3. This covers Siemens and Philips scanners, and 1.5 T and 3 T scanners.

Chapter 8 gives the evaluation of these sequences, in healthy volunteers. Chapter 9 then provide an evaluation in "tumours": firstly, in a phantom containing solutions in

small spheres to represent tumours, and then in a small cohort of patients with HCC. The method to create the phantom is described in detail, as its design is novel.

Chapter 10 concludes the thesis, summarising the project and outlining future work needed in this area.

2. Hepatocellular Carcinoma and Iron Sparing

This chapter provides the biological and clinical context regarding HCC and the need for earlier detection. It outlines why measurement of iron in HCC is of interest and how detecting iron sparing could be useful as a tool for early diagnosis.

2.1. The liver

The liver is located in the right upper abdomen (Figure 1).

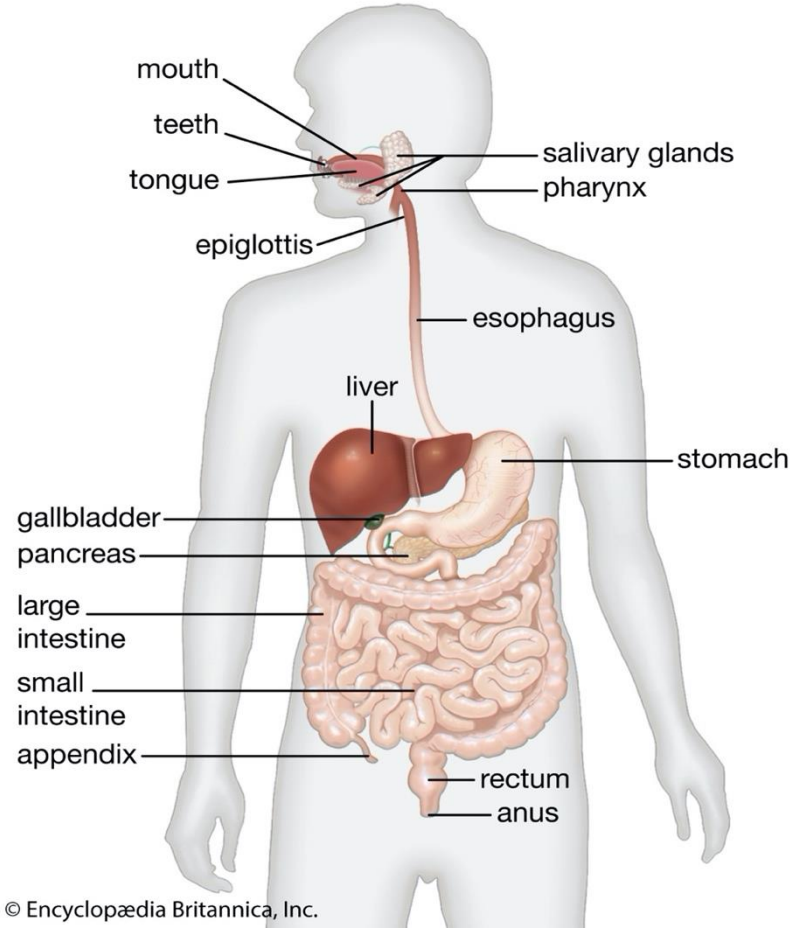


Figure 1 The human digestive system with the liver located in the right upper abdomen[23]

The liver is divided into two lobes, the left and right lobe, as shown in Figure 2. Oxygenated blood is supplied to the liver through the hepatic artery (arterial blood). Deoxygenated blood is delivered to the liver through the portal vein (venous blood). In

images of the liver the aorta and inferior vena cava may be seen (see Figure 2), however, these are not connected directly to the liver. Blood flows through the liver in sinusoids (Figure 3). Blood leaves the liver through the hepatic vein (see Figure 2 and Figure 3).

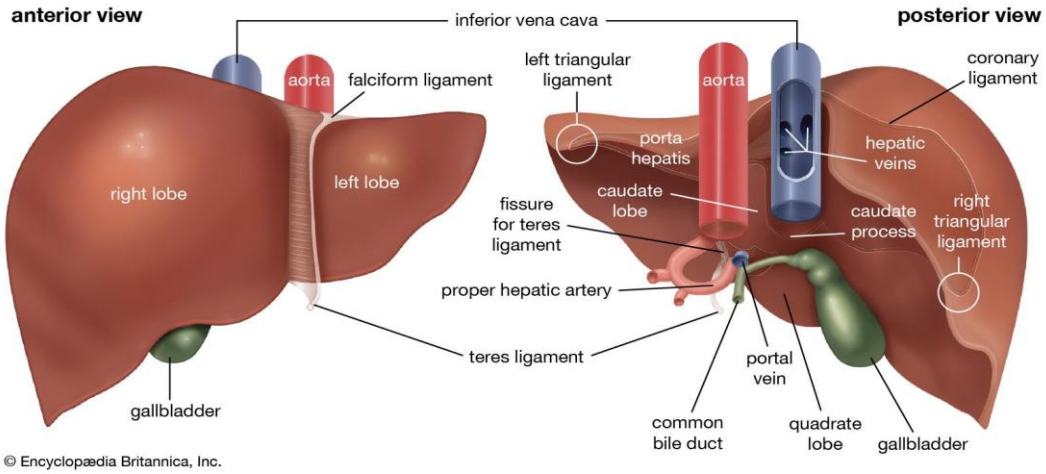


Figure 2 Structure of the liver, left and right lobes, hepatic artery and portal vein[24]

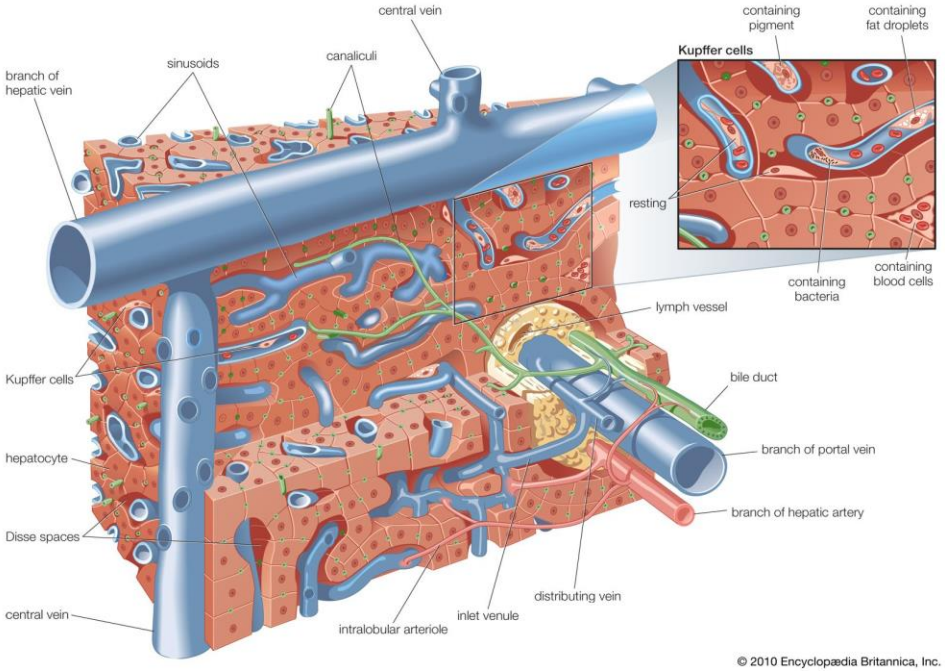


Figure 3 Microscopic structure of the liver, showing sinusoids, hepatocytes and Kupffer cells[25].

Most of the liver parenchyma (tissue) is made of liver cells called hepatocytes which perform many metabolic functions. Kupffer cells line the sinusoids and perform important functions including ingestion of foreign particles and production of anti-bodies.

As mentioned in Chapter 1, functions of the liver include the removal of toxins from blood, bile production (needed for the absorption of fats), albumin production (a protein required to maintain the correct pressure to keep fluid in blood vessels, as well as to transport substances around the body) and storage of vitamins and minerals[1]. Once produced, bile travels through the liver in ducts, before leaving through the common hepatic duct. Bile is then transported to either the gall bladder or duodenum. The gall bladder can sometimes be seen on images of the liver, such as MR images. Another feature that can be seen on images is the porta hepatis, which is a fissure in the liver where vessels enter or leave the liver.

Medical images of the liver are typically viewed in the axial, coronal and sagittal view. Figure 4 shows the different planes and the MRI images in each plane.

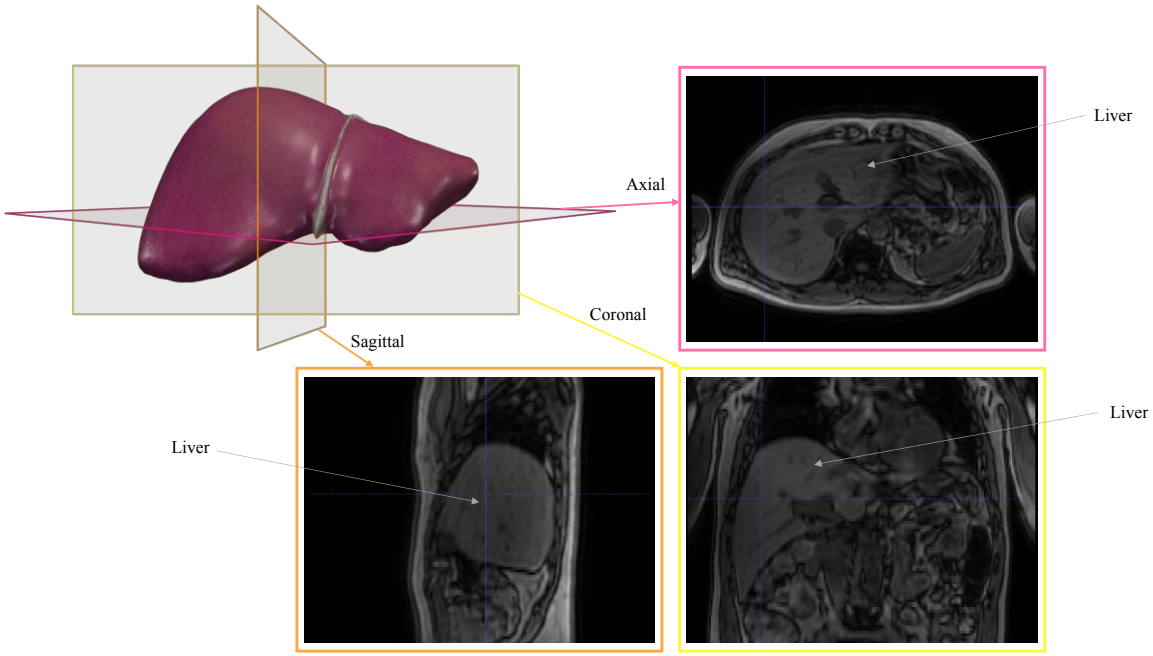


Figure 4 A diagram of the liver with axial, coronal and sagittal planes, with MRI images of the liver in each orientation

2.2. Cirrhosis

As mentioned in Chapter 1, 70 – 90% of HCC occurs on the background of liver cirrhosis[5]. Cirrhosis develops after repeated hepatic insult which may be due to viral hepatitis, excess alcohol consumption, or non-alcoholic steatohepatitis. Insult activates hepatic stellate cells which produce collagen and glycoproteins[26]. At the same time, collagen breakdown is impaired. This results in the build-up of the extracellular matrix (non-functional extracellular tissue)[26]. The accumulation of this scarred matrix is called fibrosis. "Cirrhosis" refers to highly advanced fibrosis, causing changes to the liver architecture. Collagen levels are substantially increased: in one study collagen in mg/g wet liver tissue was four to seven fold higher in cirrhotic compared to healthy livers[27]. Cirrhosis was previously considered as end-stage and irreversible, but there is increasing evidence for regression of fibrosis in cirrhotic patients where their chronic liver disease has been treated[28].

The 1977 WHO definition of cirrhosis was that it is "a diffuse process characterized by fibrosis and the conversion of normal liver architecture into structurally abnormal nodules."[3] As well as diffuse fibrosis and nodules there may be altered lobular architecture, vascular thrombosis and obliterative lesions in portal tracts and hepatic veins and under-perfusion of lobular parenchyma and consequent tissue hypoxia[29]. It has been argued that the WHO definition has limitations. The International Liver Pathology Study Group has challenged its usefulness and recommends that is no longer used[28]. This is because cirrhosis is a heterogeneous condition with different clinical symptoms and prognosis, depending on aetiology and severity of hepatic architectural distortion, therefore it has limited clinical utility. Despite its limitations and broad scope, cirrhosis is still the term used in clinics throughout the world. It is diagnosed primarily using biopsy, based on late stage evidence of fibrosis (in Pathology this is METAVIR fibrosis

stage 4, or Ishak stages 5 and 6)[29], although liver stiffness, radiology, blood markers and/or clinical symptoms can sometime be used instead. Commonly, it is split into two categories: compensated and decompensated cirrhosis. Compensated cirrhosis is when the liver still can carry out most functions, whereas decompensated cirrhosis is when the liver becomes unable to function adequately and clinical symptoms become manifest, in particular clinically evident complications of portal hypertension[26 30]. However, even within these categories it has been argued that further refinement is required[30].

2.3. Hepatocarcinogenesis

Hepatocarcinogenesis (the process of the development of HCC) is a complex process. HCC often occurs on the background of cirrhosis, but it is not clear whether cirrhosis causes HCC or whether they develop largely separately. Generally, HCC develops more slowly than cirrhosis, and so appears later. It might be that cirrhosis and hepatocarcinogenesis are separate parallel processes, both caused by inflammation[6]. The mechanisms may also depend on the underlying liver disease e.g. hepatitis B infection or alcoholic liver disease[31]. Mechanisms proposed include cycles of necrosis, inflammation and regeneration, oxidative stress, viral integration into the host genome, inactivation of the p53 tumour suppressor, genomic instability and genomic alterations[31]. Globally, the most common underlying liver diseases in HCC patients are: hepatitis B (~50% of HCC cases); and hepatitis C (~25% of HCC cases)[5]. However, HCC can also develop on the background of other underlying diseases such as alcoholic liver disease, non-alcoholic steatohepatitis, and haemochromatosis. Table 1 gives the HCC incidence for patients with different types of liver disease with cirrhosis.

Underlying disease	HCC incidence per year (%)
Chronic hepatitis C	1-8
Chronic hepatitis B	1-15
Alcoholic liver disease	1
Non-alcoholic steatohepatitis	2.6
Hemochromatosis	2-6
Primary biliary cirrhosis (PBC)	2
Autoimmune hepatitis	< 0.2

Table 1 HCC incidence per year in patients with cirrhosis with different underlying liver diseases[5 32 33]

Although less common, HCC can also develop in non-cirrhotic livers, which also supports the proposition that cirrhosis and hepatocarcinogenesis may be separate, parallel processes. One example of HCC developing in non-cirrhotic livers is in patients with non-alcoholic fatty liver disease (NAFLD). Several studies have reported that 40-50% of patients with HCC and NAFLD have non-cirrhotic livers[34]. As with the development of HCC in other diseases on the background of cirrhosis, a variety of similar mechanisms have been suggested such as inflammatory response and oxidative stress[34]. Another example of HCC developing in non-cirrhotic livers is its occurrence in hepatitis B patients without cirrhosis (although 70-80% of hepatitis B patients with HCC do have cirrhosis)[35]. The mechanism behind HCC development in patients with hepatitis B is not fully understood, but it is thought that the virus itself may directly cause HCC[36].

At a molecular level, hepatocarcinogenesis may be sub-divided into multi-step and *de novo*. In multi-step hepatocarcinogenesis, nodules progress from benign to cancerous. A more aggressive sub-nodule grows within a more benign nodule and eventually replaces it[37]. It has been hypothesised that the pathway might be as follows: Cirrhotic Nodule – Low Grade Dysplastic Nodule – High Grade Dysplastic Nodule – Early HCC – Progressed HCC. Not all steps may occur; some may be skipped. The multi-step process may occur simultaneously in different areas of the liver, and at different speeds[37]. In *de novo* hepatocarcinogenesis, HCC arises with no (apparent) previous lesions[6]. It should

be noted that molecular hepatocarcinogenesis is always multi-step, but not at a histopathological level[6].

During hepatocarcinogenesis, various pathophysiological changes take place, one of which is the change in blood flow of the tumour compared to the liver parenchyma. This is depicted in Figure 5. The blood flow from the hepatic artery and portal vein to a nodule changes during carcinogenesis. Portal tracts (also known as portal triads) consist of branches of the (non-tumoral) hepatic artery and portal vein, as well as bile ducts. During hepatocarcinogenesis, blood supply from the portal tract decreases due to stromal invasion and other factors[6]. However, unpaired arteries (i.e. arteries unaccompanied by portal veins or bile ducts) develop due to angiogenic factors[6], which increase the blood supply. Consequently, dysplastic nodules and early HCCs generally have decreased arterial flow, followed by increasing arterial flow in progressed HCCs. Infiltrative HCCs can have a decrease in arterial flow.

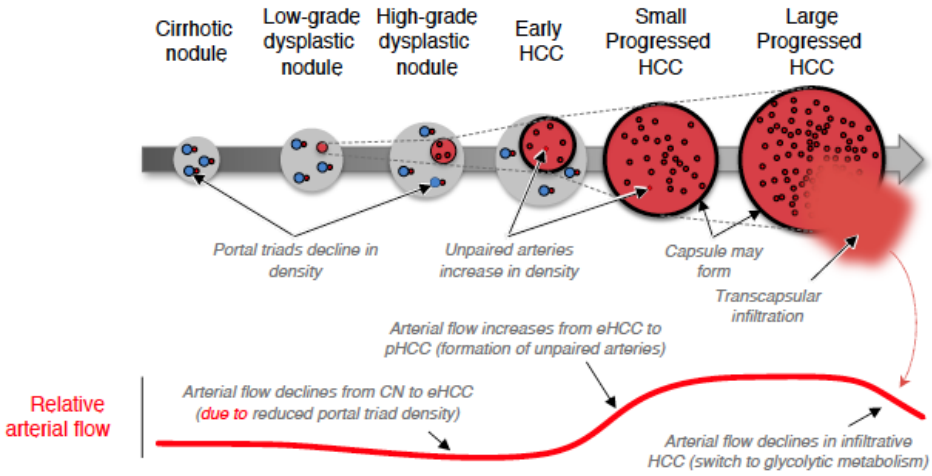


Figure 5 Alteration to arterial flow during hepatocarcinogenesis, Source: LI-RADS v2018[6]

2.4. Imaging of HCC

As mentioned in Chapter 1, surveillance for HCC in at-risk populations on a six monthly basis using ultrasound is a strategy used and recommended by medical associations and societies throughout the world[9–11].

Current guidelines recommend multiphase contrast enhanced CT or MRI for diagnosis[6 9 10]. The Liver Imaging Reporting and Data Standards (LI-RADS) from the American College of Radiologists provides specific guidance on imaging and its interpretation for diagnosis of HCC. LI-RADS aims "To standardize the terminology, technique, interpretation, reporting, and data collection of liver imaging for clinical care, education, and research in patients at high risk for or with HCC"[6]. LI-RADS classifies lesions depending on image features, with LR-1 being definitely benign and LR-5 being definitely HCC. A combination of features during the different phases enables this categorisation. It should be noted that LI-RADs only applies to patients at high risk for HCC (i.e., with cirrhosis, chronic hepatitis B or current/prior HCC).

LI-RADS primarily involves the use of contrast agents to perform the diagnosis of HCC. The underlying physics of the use of contrast agents in MRI is given in Section 3.4, but the basic principle is that as they travel through the liver with the blood the differences in vascularity between the tumour and parenchyma cause the tumour to show different contrast to the parenchyma, and hence enable diagnosis. Traditionally contrast agents are extracellular, that is they are distributed in the blood and extracellular space in the body, and it is extracellular agents which LI-RADS is primarily based on. According to LI-RADS the features of confirmed HCC (LR-5) are:

- For lesions 20mm or more: non-rim arterial phase hyperenhancement and one or more of enhancing capsule, non-peripheral washout, and threshold growth
- For lesions 10-19mm: non-rim arterial phase hyperenhancement and two of enhancing capsule, non-peripheral washout, and threshold growth, or one of non-peripheral washout and threshold growth.

LI-RADS also gives optional ancillary features which favour malignancy (either in general or HCC specific). These can be used to upgrade LR-1 to LR-3 lesions to higher LR-scores, but they cannot be used to upgrade LR-4 to LR-5.

As well as extracellular agents, hepatobiliary agents/liver-specific agents are available such as Primovist. For Primovist imaging, like extracellular agent imaging, there are arterial and portal venous phases, then instead of a delayed phase there is a transitional phase, as agent begins to be taken up by the hepatocytes. This is followed by the hepatobiliary phase, 10-20 minutes after contrast injection, where the contrast has entirely left the vessels and has been taken up by the hepatocytes. About 50% of Primovist is excreted through the bile (after being transported through the hepatocytes), and 50% is excreted through the kidneys, via the blood[6]. This is different to extracellular contrast agents, which are excreted entirely through the kidneys.

Hepatocellular function of HCC declines during hepatocarcinogenesis[38]. Of particular importance for Primovist imaging, organic anionic transporting polypeptides (OATPs) expression decreases. OATPs are a family of proteins located on the sinusoidal membrane, and transport bile salts into the hepatocytes. The result is that, even for early HCCs, there is no or very little uptake of Primovist by the tumour during the hepatobiliary phase and so the tumour appears hypointense, in comparison to the surrounding liver. In LI-RADS hypointensity on the hepatobiliary phase is classed as an ancillary feature favouring malignancy i.e. an additional feature which may aid with detection and diagnosis of HCC[6]. Use of the hepatobiliary phase can improve per-lesion sensitivity by 6-15% compared to gadoxetic acid[38]. It is especially useful for detecting early HCCs, as OATP expression declines prior to complete neo-arterialization, so an early HCC may be not be enhanced on the arterial phase, but may be hypointense in the hepatobiliary phase[38]. However, 5-10% of HCCs are hyperintense during the hepatobiliary phase[6].

2.5. HCC Staging and Treatment

Both EASL and AASLD guidelines endorse the Barcelona Clinic Liver Cancer (BCLC) staging system[9] [10]. This staging system is based on imaging and clinical criteria, which involve a measure of liver disease called Child Pugh, and performance status (a measure of the patient's ability to carry out daily life activities). The BCLC stages and the recommended treatment from EASL are shown in Table 2.

Stage	Imaging Criteria	Clinical Criteria	Treatment (as recommended by EASL)
Very early (0)	1 HCC 1-2cm	Child Pugh A, Performance Status 0	Ablation/Resection
Early (A)	1 HCC 2-5cm OR 2-3 HCCs 1-3cm	Child Pugh A-B, Performance Status 0	Ablation/Resection/Transplant
Intermediate (B)	1 HCC >5cm OR 2-3 HCCs >3cm OR 4 or more HCCs of any size without tumour in vein	Child Pugh A-B, Performance Status 0	Chemoembolization
Advanced (C)	Tumour in vein regardless of size or number of HCCs OR Nodal metastases OR Extra-hepatic metastases	Child Pugh A-B, Performance Status 1-2	Systemic treatment
End Stage (D)	Stage not dependent on imaging criteria	Child Pugh C, Performance Status 3-4	Best supportive care

Table 2 BCLC stages[6] [9]

Outcomes depend on the stage when the tumour is diagnosed. If HCC is detected at Stage 0 (very early) survival rates following resection are 80-90 %, and these rates are similar for radiofrequency ablation[9]. If detected at Stage A (early) then median 5-year survival rates following treatment (ablation/resection/transplant) are 50-70%[9]. This is in contrast with outcomes if patients are diagnosed at Stage C (advanced), where the median survival rate after one year is just 25%[9], although progress has been made with the development of drugs such as sorafenib, regorafenib and lenvatinib. These differences in survival rates emphasise the need to detect HCC as early as possible, when treatment leads to good survival rates.

2.6. Iron and Hepatocellular Carcinoma

2.6.1. Iron in the liver

The liver helps maintain iron homeostasis in the body. It performs a variety of functions to do this including being the main storage site for iron and the main site for synthesis of the hormone hepcidin, which regulates iron[39]. All liver cells can store iron, but in normal conditions the majority of iron is stored in hepatocytes[39]. Iron is generally stored in the protein ferritin, but can be stored in hemosiderin, an iron-storage complex, if ferritin stores become saturated[39]. Disorders such as hereditary haemochromatosis, a genetic iron overload disorder, lead to dysregulation of the liver's function in maintaining iron homeostasis and can result in iron overload, both in the liver and the body[39]. In diseases such as thalassemia, iron can also be increased due to treatment involving multiple blood transfusions[40]. In these patients monitoring of liver iron is important.

2.6.2. Iron in the liver parenchyma and HCC development

Increased iron is concerning as it increases the risk of liver cancer. Various mechanisms have been proposed for this, with one key mechanism being the role that iron plays as a catalyst in the generation of reactive oxygen species (ROS)[41]. This elicits toxic effects and promotes carcinogenesis[41]. Excess iron in the liver can therefore promote carcinogenesis directly. The toxic effect of iron on the liver can also promote liver injury and cirrhosis, and thereby promote carcinogenesis indirectly. Patients with hereditary hemochromatosis have a 20-200 fold increased risk in developing HCC[41].

Iron may also act synergistically to promote carcinogenesis in other disease types, such as non-alcoholic steatohepatitis (NASH). In a study of NASH patients, 51 of which had HCC and 102 without HCC, an iron score (0-60) based on biopsy was used to compare iron in the liver parenchyma between the two groups[42]. They found that hepatic iron was significantly higher in patients with HCC than in HCC-free controls[42]. However, the iron

found in NASH patients is well below the thresholds associated with increased risk of HCC in hereditary haemochromatosis patients[43], so in these cases it is a mild increase in iron compared to normal levels which may be important to measure[43].

2.6.3. Iron in HCC

Although increased iron in the liver parenchyma is associated with increased risk of HCC, several studies suggest that as HCC develops the iron in the pre-cancerous/cancerous nodule decreases[17–19], which is thought to be due to clonal expansion of malignant cells which have iron resistance[37]. When there is less iron in the nodule than in the surrounding liver it is called iron sparing.

Studies investigating the occurrence of iron sparing in HCC involved sampling liver and tumour by biopsy and measuring the iron concentration of each. Iron concentration can be either measured as dry weight or wet weight iron concentration.

One study where dry weight (dw) iron concentration was measured was by Ebara *et al.*[17] which included 113 patients with HCC. In this study they found mean +/- standard deviation (SD) of HCCs was 1.3 +/- 0.1 mg/g dw vs 1.7 +/- 0.1 mg/g dw in liver tissue[17]. Using the Wilcoxon signed-rank test the difference between HCC and liver tissue was found to be statistically significant. A study by Tashiro *et al.*[19] of 23 patients with HCC found a statistically significant difference in iron concentration between HCC and liver (the statistical test was not specified in their paper)[19]. Unlike other studies evaluating iron in the liver and tumour they reported the geometric mean and SD as they found that trace elements in the liver had a log-normal distribution in a different study[19]. As no arithmetic mean and SD was reported, it is therefore difficult to directly compare their results with other studies. They found the geometric mean was 0.4 mg/g dw with geometric SD of 2.5 for HCCs, compared to the liver the geometric mean was 0.6 mg/g dw with geometric SD of 3.1[19].

A study by Honda *et al.*[18] of 35 patients using wet weight (ww) concentrations again found a statistically significant difference between HCC and liver parenchyma iron concentration using Wilcoxon signed-rank test. In this study the mean +/- SD iron concentration was 0.09 +/- 0.05 mg/g ww in HCC vs 0.1 +/-0.09 mg/g ww in liver[18]. A figure from this paper is shown in Figure 6. As can be seen on the figure, the difference in iron concentration between HCC and liver varies between patients, some with much larger differences than others and some where there was higher iron in the HCC compared to the liver.

Figure removed due to not receiving permission from authors to use their figure in this thesis. Please refer to Figure 3 in Reference 18 to see this figure.

Figure 6 Comparison between iron concentration (μg/g ww) between liver and HCC in 35 patients from Honda *et al.* study, copied from their paper[18].

In another study by Honda *et al.*[21] of 13 patients with HCC, in the HCC mean +/- SD was 0.04 +/- 0.03 mg/g ww in the tumour vs 0.06 +/- 0.04 mg/g ww in liver parenchyma. The difference in iron concentration between HCC and liver was not statistically significant, however they used Student's paired t-test, which might not have been appropriate if the data was not normally distributed. Another study by Ebara *et al.*[20] of 59

patients also did not find statistical significance between HCC and liver, when using Wilcoxon signed-rank test. The mean +/- SD for HCC was 1.4 +/- 1.1 mg/g dw in HCC vs 1.4 +/- 0.9 mg/g dw in the liver. It is unclear why this result is different from the studies mentioned earlier which did show a statistically significant difference in iron between HCC and liver. It may be due to differences in aetiology, for example. This study may give evidence that iron sparing does not always occur in HCC.

A further study to support that a feature of HCC is iron sparing, compared to dysplastic nodules which are not cancerous, is to be found in a study by Li *et al.*[44]. This study did not involve measurement of iron concentration but instead used iron scores, based on biopsies with Prussian Blue staining (score 0-4 from no siderosis to severe siderosis, siderosis is excess iron deposition). All 12 dysplastic nodules contained iron and there was no statistically significant difference between the nodule and liver (calculated using a paired t-test). For the 64 HCCs, 95.3 % contained no iron, and there was a statistically significant difference between HCC and liver[44].

The limited number of studies and differences in conclusions indicate that although iron sparing does seem to be a feature of HCC, more data is required to verify this. If iron sparing is a feature, then it could potentially be used as a way of detecting early HCC. The limited number of studies is probably due to a variety of factors including the requirement in the past to have biopsy to measure iron. Not all patients with HCC have biopsy: diagnosis can be performed using imaging alone (see LI-RADS[6]). Even if a biopsy is collected, measuring iron concentration is not routine clinical practice in HCC patients, particularly as it generally involves destroying the sample (atomic absorption spectrophotometry is often used). This means that multiple biopsies of tumour and parenchyma need to be taken. Staining can be used but this only leads to a score of iron, not a measure of concentration.

A non-invasive tool is critical if the literature is to become more precise regarding iron sparing during hepatocarcinogenesis and thus provide verification if it could be a useful feature for early detection of HCC.

2.7. Summary

In summary, there is an unmet need to detect HCC as early as possible when curative treatment is possible. Iron sparing may be a critical, clinically significant, stage in the development of HCC, but discussions in the literature are undermined by the lack of a sufficiently accurate method to detect and measure iron in the liver and tumour non-invasively.

3. MRI Physics and Detection of Iron using MRI

3.1. Introduction

This chapter outlines some basic MRI physics, which is essential to understanding the work of this thesis, and then the use of MRI to detect iron sparing is discussed.

3.2. Signal Generation

MRI involves placing a participant within a static magnetic field, called B_0. The static magnetic field is usually measured in tesla (T). Tesla is a measure of magnetic flux density. In clinical practice this is typically of strength 1.5 T or 3 T. For comparison, the magnitude of the Earth's magnetic field at the Earth's surface is around 0.3 x 10^{-4} T to 0.6 x 10^{-4} T.

The body is made up of around 60% water (H_2O). The nucleus of the hydrogen (H) atoms in water consists of a single positively charged proton. Protons have an intrinsic quantum property called spin angular momentum, or spin. This gives rise to the proton having a magnetic moment. The magnetic moment is given by $\mu = \gamma S$, where γ is the gyromagnetic ratio (for hydrogen 1H this is 42.58 MHz/T) and S is the spin angular momentum. When placed in an external magnetic field, B_0, the magnetic moments will either orientate themselves towards or against the external field. Fractionally more protons (on the order of parts per million) will align *with* the field. The vector sum of the magnetic fields of the protons in tissue will thus be non-zero and is called the net magnetization (M_0).

The protons also precess around the axis of the magnetic field. The rate of precession is given by the Larmor equation $\omega = \gamma B_0$, ω is the angular precessional frequency, γ is the gyromagnetic ratio (for hydrogen 1H is this 42.58 MHz/T) and B_0 is the external

magnetic field strength. The B_0 field direction is often assigned to the z direction. The conventional axes used in MRI are shown in Figure 7, with the B_0 field direction.

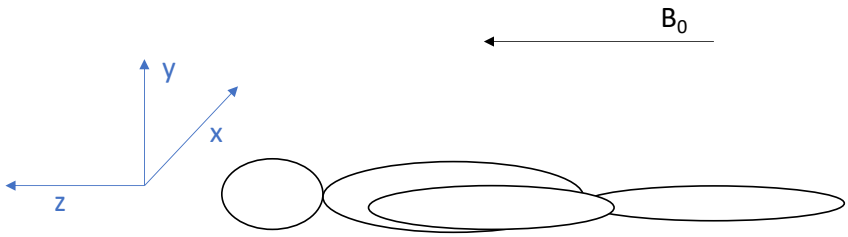


Figure 7 Conventional axes assignment in MRI, with the magnetic field, B_0, along the z direction, typically the "head" direction as indicated by the human form, although patients may be oriented feet first as well.

To generate a measurable signal, a radiofrequency (RF) pulse with the Larmor frequency is applied, perpendicular to the B_0 field. This oscillating magnetic field is called B_1. The protons will now precess around the new field, as well as well as B_0, resulting in a spiral rotation of M_0 towards the x-y plane. The amount to which M_0 flips towards the x-y plane is called the flip angle and is proportional to the strength and duration of the RF pulse. The protons also will now precess in phase, until the RF pulse is turned off. Magnetization in the x-y plane is called transverse magnetization, whereas in the z direction it is called longitudinal magnetization.

The rotating magnetic field in the transverse plane induces a current varying at the Larmor frequency in the detector coils, which produces a signal with a sine wave oscillating at the Larmor frequency. When the RF pulse is turned off, the protons become increasingly out of phase, due to B_0 inhomogeneities and tissue-specific spin-spin interactions. This results in decay of magnetization in the transverse plane, which is termed free induction decay (FID) and is shown in Figure 8. This exponential decay has time constant T_2*. However, in MRI the FID is not measured directly, instead, "echoes" are used, and the two that are used most in current clinical practice are the gradient echo and spin echo.

Gradients are small changes in the magnetic field in space and are created using gradient coils. In gradient echoes, a negative gradient is applied in the x direction, which causes rapid dephasing (even faster than FID). A positive gradient is then applied in the x direction, which rephases the protons, creating an echo. The time from the centre of the RF pulse to the centre of the echo is called time to echo (TE). Time between repeated pulses is called the repetition time (TR). The amplitude of the signal is determined by T_2*. Gradient echoes are often referred to gradient recalled echoes (GRE). The signal during a GRE is shown in Figure 8.

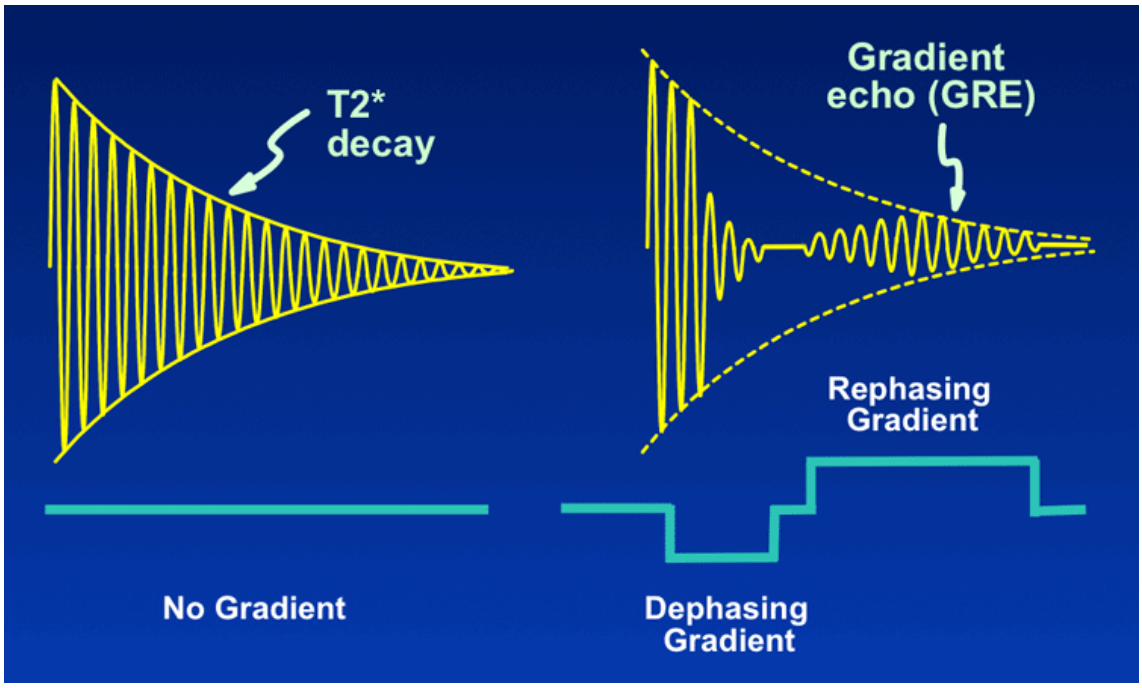


Figure 8 Signal during free induction decay (FID) (left), labelled T2* decay, and signal from a GRE sequence (right) Courtesy of Allen D. Elster, MRIquestions.com

Alternatively, in the case of spin echoes (SE), after an initial RF pulse (typically 90°), another pulse (typically 180°) is applied. This pulse reverses the phase angles of the spins and brings them back into phase. This rephasing pulse is applied at TE/2 relative to the first pulse. The result of applying the second pulse is that dephasing due to B_0 inhomogeneities is counteracted, and so the amplitude of the signal is only determined by the decay due to tissue-specific spin-spin interactions, which is called T_2 decay (although other physiological processes may also contribute, such as diffusion and motion). The

exponential decay has time constant T_2. Figure 9 shows the signal during a spin echo sequence.

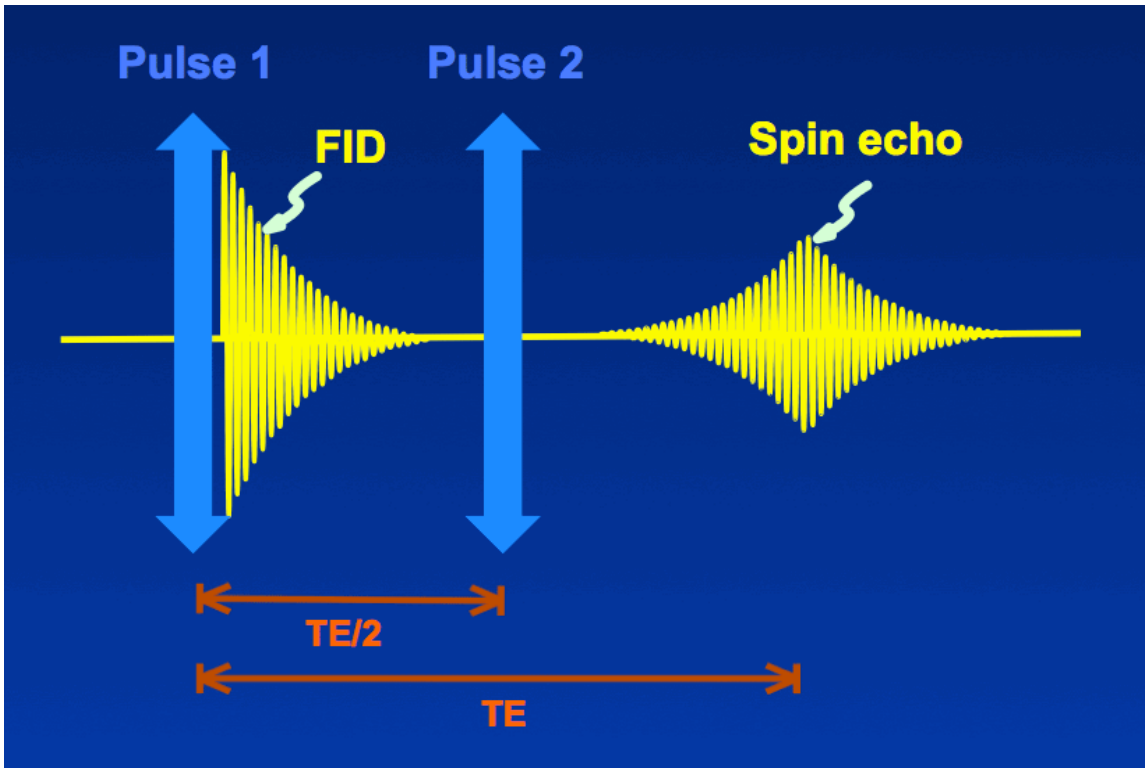


Figure 9 Signal during a spin echo. Pulse 2 occurs at TE/2 relative to Pulse 1. TE = echo time. Courtesy of Allen D. Elster, MRIquestions.com

As well as T_2 relaxation, T_1 relaxation also occurs. This involves the growth of longitudinal magnetisation back to the original equilibrium state and occurs due to tissue-specific spin-lattice interactions. This recovery of longitudinal magnetization takes longer than the decay of transverse magnetization and is a separate process. T_1 is the time constant of the exponential growth of the signal. A diagram showing T_1 and T_2 relaxation is given in Figure 10.

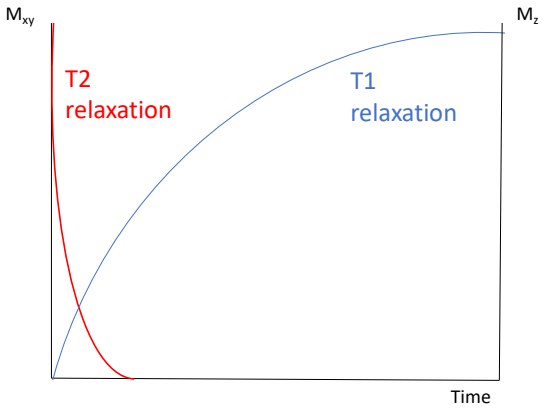


Figure 10 T_1 and T_2 relaxation curves: T_2 relaxation occurs much faster than T_1 relaxation.

3.3. Spatial encoding

To create an image in MRI, spatial encoding is required. To this end gradients are used. When gradients are applied the magnetic field at each location along an axis is changed from B_0 to $B_0 + dB_x$, where x is the spatial coordinate. In frequency encoding the gradient is applied during the read out of the signal, which changes the Lamour frequency as a function of position. It is the frequency encoding gradient which is used to generate gradient echoes. Frequency encoding could be performed in all three directions. However, applying spatial gradients in all three directions (x, y and z) during the signal acquisition would result in multiple voxels having the same frequency and so does not allow for spatial encoding of the image. To construct an image phase encoding is used in one direction (e.g., the y direction), which involves repeating the acquisition multiple times, each with a different gradient, and so during each repetition the field strength changes as a function of y i.e., $B_0 + dB_y$, where y is a different location in the y direction. MRI imaging can be 2D or 3D. If 2D, then only a slice of the object is imaged. To do this a slice select gradient is applied in the z direction at the same time as the radiofrequency (RF) pulse. The RF pulse will have a bandwidth such that only protons within the desired slice will have frequencies such that resonance will occur. For 3D MRI imaging, instead

of having a slice select gradient in the z direction, phase encoding is applied in both the z and y directions.

Applying gradients does not immediately result in an image: instead, a data space called k-space is filled. An inverse Fourier Transform can then be applied to the k-space data to produce an image. The most common method to fill k-space is to use Cartesian sampling. During readout of the signal, a row in k-space is filled. The frequency encoding gradient changes the frequency during the readout. At the beginning and end of the sampling the gradient is the most negative/positive, and higher spatial frequencies are captured, whereas towards the middle of the sampling lower spatial frequencies are sampled. Each row in k-space corresponds to a different gradient strength in the y direction: the sequence is repeated with different gradient strengths in the y direction (negative through to positive), prior to sampling the data (phase encoding). The strongest gradients are in the top and bottom rows, and these correspond to the highest spatial frequencies in the y direction. In the centre where no/little gradient is applied corresponds to lower spatial frequencies.

An example of how k-space corresponds to different spatial frequencies in the final image is shown in Figure 11. In 2D MRI, k-space is 2-dimensional. In 3D MRI, k-space is 3-dimensional, and a 3D inverse Fourier Transform is applied to construct the 3D image.

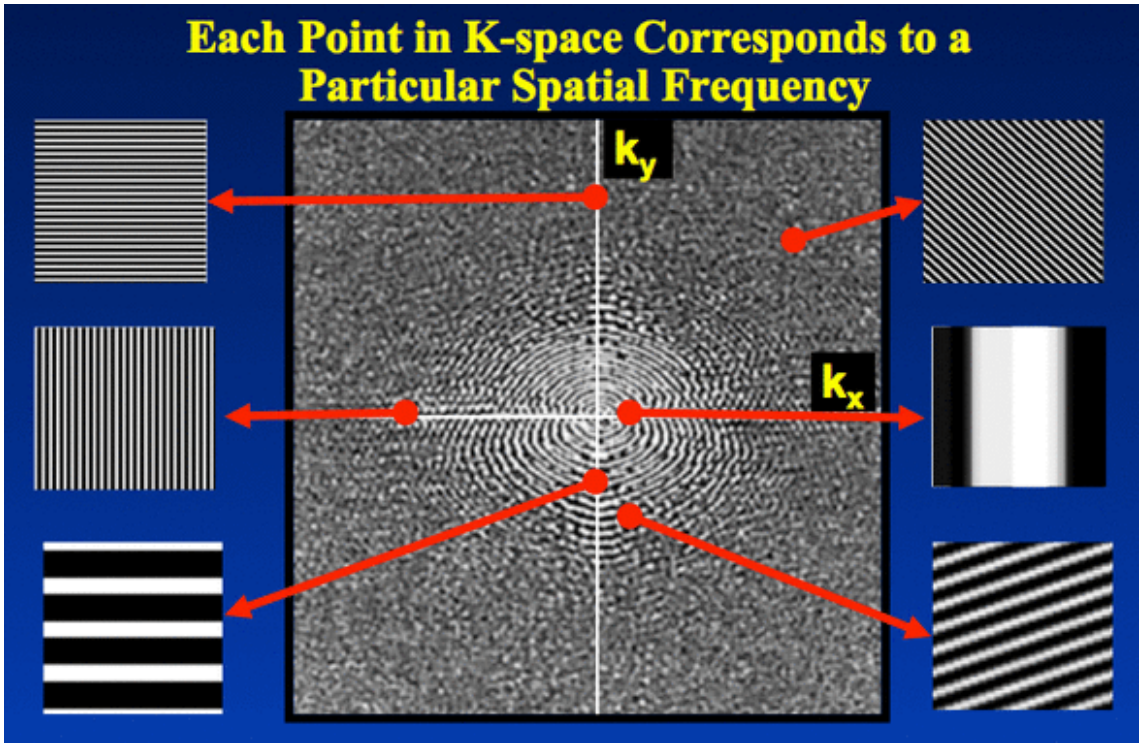


Figure 11 Points in k space correspond to different spatial frequencies: those at the edge correspond to higher spatial frequencies. Those further from the central axes correspond to spatial frequencies at a higher angle to the horizontal/vertical. Courtesy of Allen D. Elster, MRIquestions.com

An MRI sequence can be depicted in a pulse sequence diagram. A simple GRE 3D pulse sequence diagram is shown in Figure 12. The need for multiple repetitions of the sequence for each phase encoding step in the y and z directions lengthens the acquisition time.

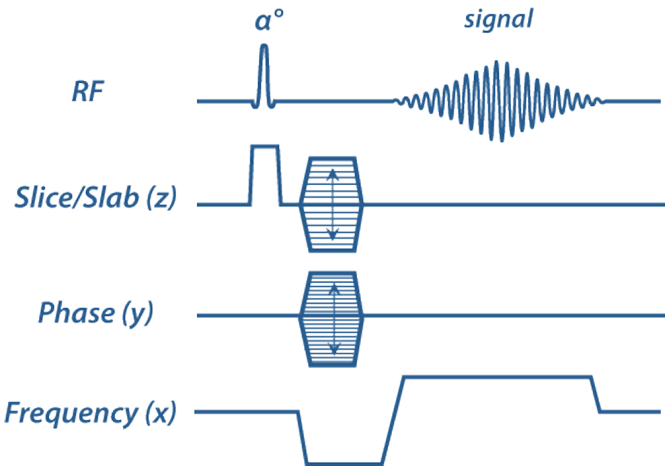


Figure 12 3D MRI GRE pulse sequence diagram. Courtesy of Allen D. Elster, MRIquestions.com

3.4. MRI Contrasts

In clinical MRI, the two main image types produced are T_1 and T_2 weighted imaging. These provide different contrasts between tissues, which allow clinicians to identify different radiological features. Differences in spin-lattice interactions and spin-spin interactions in different tissues produces the contrast between tissues in T_1-weighted images and T_2-weighted images respectively. Choice of TR and TE can be used to alter the weighting of the image. Longer TEs mean that there has been longer for T_2 decay at the time of the echo and so differences in T_2 decay between tissues are more likely to be apparent. Short TRs are used for T_1-weighted imaging: if the TR is too long then all longitudinal magnetization will have recovered before the next pulse, whereas if it is short tissues will have experienced different amounts of T_1 relaxation, and so the image will have T_1-weighting. Therefore, long TEs and TRs are used for T_2-weighted images and short TEs and TRs are used for T_1-weighted images. Both SE and GRE sequences can be used for T_1 weighted images, but only SE sequences can be used for T_2-weighted images. GRE sequences can also be used for T_2* weighted images.

Contrast agents can also be used to alter the contrast between tissues. Most MRI contrast agents are gadolinium chelates. Gadolinium is strongly paramagnetic and shortens the relaxation time of surrounding hydrogen protons. This results in tissues containing contrast agent being bright on T_1-weighted images. In MRI of HCC, the presence of gadolinium in the arterial blood causes the tumour to be bright during the arterial phase on T1-weighted images. Primovist, a liver-specific contrast agent, contains gadolinium but has a lipophilic EOB group which extracellular agents do not have. This means that the agent can be taken up by hepatocytes if OATP are present, and the presence of the gadolinium will be shown by that area being bright on T_1-weighted imaging. Thus in HCC, which often have decreased OATP expression, the tumour appears hypointense,

as contrast agent is not taken up in the tumour, but is taken up by the hepatocytes in the liver parenchyma[6].

3.5. Acceleration in MRI

Acquiring data in MRI is a slow process, primarily due to the need for multiple repetitions of the pulse sequence to allow for spatial encoding in the y direction (and z if in 3D). The higher the resolution or spatial coverage, the more repetitions are needed. Over the years various techniques have been proposed to accelerate MRI. A few of these techniques are covered in this section as these are relevant to later work in this thesis.

3.5.1. Partial Fourier

In theory, k-space has conjugate symmetry. This means that it should be possible to only sample half of k-space and estimate the other half, thus cutting the number of phase encoding steps by half and reducing the scan time. However, the assumption that k-space is symmetric is dependent on the underlying B_0 field being the same everywhere. However, this is unlikely to be a valid assumption as there are often variations in the B_0 field and localised susceptibility effects. To enable a better estimation of the k-space which has not been sampled, over half of k-space can be sampled (typically 60% or more) and the centre portion of k space used to create phase correction maps. This type of sub-sampling of k-space with estimation of the remaining portion using conjugate symmetry is called Partial Fourier and is a very commonly used technique on clinical scanners. It is primarily used in the phase encoding direction(s) (it can include the slice direction in 3D imaging). The scan time is reduced by the same percentage decrease as the percentage decrease of the number of measured phase encoding steps, so a 40% reduction in phase encoding steps, would result in 40% reduction in scan time. Hence, acceleration using Partial Fourier has maximum acceleration is around 1.7 times.

Using Partial Fourier has a number of limitations. Firstly, higher frequency information is only sampled once in the phase encoding direction and so this may result in less accurate delineation of small structures such as vessels. Secondly, signal to noise ratio (SNR) is decreased due to fewer measured phase encoding steps. Thirdly, if there are large phase shifts these may not be corrected for properly, for example at air-tissue interfaces.

3.5.2. Parallel Imaging

Another way to reduce scan time is to omit phase encoding steps, for example, omit every second phase encoding step. However, this results in an aliased image with a reduced field of view as the sampling is not sufficient to represent the object.

Originally, single receiver coils were used in MRI, but using multiple surface coils improves signal to noise ratio and produces better images. Following the development of multiple surface coils, a further major development in MRI was made with the introduction of parallel imaging. In parallel imaging the presence of multiple receiver coils is used to “unwrap” aliased images which result from under sampling.

One way to do this is in the image domain. In this method coil sensitivity maps are generated using a calibration scan. The acquired (aliased) signal is the sum of the “real” pixel values (at two locations) multiplied by the coil sensitivity (at those two locations). Using multiple coils leads to multiple equations of this format, which then can be solved to give the “real” pixel values at each location. This method is called SENSitivity Encoding (SENSE). SENSE is used on Philips scanners. The amount of acceleration (and thus sub-sampling), also known as the reduction factor, is determined by the SENSE factor, which the user can define. A factor of 2 would indicate a 2 times acceleration, so the scan time would be reduced by half. The SENSE factor does not need to be an integer value.

Another way of performing parallel imaging is to generate the missing phase encoding lines in k-space. To do this the centre of k-space is fully sampled. This is termed the auto-calibration signal (ACS) region. Using weighting factors for each coil determined using the ACS, and the local information in k-space (in each kernel), missing k-space points are estimated for each coil. These estimations are then combined to create the final image. This technique is called GeneRalized Autocalibrating Partial Parallel Acquisition (GRAPPA) and is available on Siemens scanners. Applying parallel imaging such as SENSE and GRAPPA in two directions in 3D MRI can result in residual aliasing artifacts and amplified noise[45]. Therefore, Siemens also offer a parallel imaging technique called Controlled Aliasing in Parallel Imaging Results in Higher Acceleration (CAIPIRINHA). CAIPIRINHA uses a specific sampling pattern in k-space which minimises artifacts in 3D accelerated imaging[45]. Siemens term for their parallel imaging products is integrated Parallel Acquisition Techniques (iPAT). The iPAT factor (i.e., the acceleration factor) only be integer values, unlike the SENSE factor.

Using parallel imaging can significantly reduce scan time but can result in impaired image quality. Firstly, residual aliasing can occur [46]. This can be particularly obvious if there is a bright edge to an object. Another disadvantage is reduction in signal to noise ratio (SNR). This is partially due to the collection of fewer data points[46], and so SNR is impacted by the choice of acceleration factor. It also depends on the geometry or g-factor, The g-factor is dependent on the number of aliased replicates, and their differences in coil sensitivity, in each pixel. This in turn depends on the number of coils and their location. The g-factor varies across the image. Hence, noise distribution can be heterogeneous, and is impacted by various factors including the position, orientation, and number of surface coils. Although in theory the acceleration factor could be as high as the number of total number of phased array coil elements, parallel imaging is typically limited to

accelerating acquisitions by up to three times[47]: using acceleration factors of 3 or more can result in impaired image quality, due to residual aliasing and poor SNR[46]. Parallel imaging can be implemented alongside Partial Fourier, resulting in further reduction in acquisition time.

3.5.3. Compressed Sensing

Yet another way of further speeding up image acquisition is compressed sensing, and this has the added advantage that it does not necessarily result in a reduction in SNR. In traditional MR imaging without acceleration k-space densely samples frequency (ω_x) and phase (φ_y) and can be regarded as a function k(x,y). Compressed sensing is based on the idea that the function k(x,y) can be sampled as a much sparser set of points $\{\omega_{xi}, \varphi_{yi}: i, j\}$ and the signal at these points are used to approximate the function k over the full domain.

To do this the fewer points in k space are sampled in a random manner (incoherently), which results in a noisy image. However, pure random undersampling is impractical, as the rapid switching of gradients would result in eddy currents. Additionally, the points at the centre of k-space have the highest signal, so random weighting throughout k-space would result in reduced SNR. To overcome these limitations, the random undersampling can be performed in the phase encode plane, and variable density pseudo-random sampling schemes can be used, with more samples acquired at the centre of k-space[48,49]. The step is to de-noise the image, which depends on it being sparse. Some images are naturally sparse (i.e., the image has a few bright pixels, whilst the rest of the image is dark i.e., the pixels are lower in value), such as magnetic resonance angiography (MRA), but for many other applications the image needs to be transformed in such a way that is becomes sparse. A discrete wavelet transform can be used to do this. In this sparse representation, all pixels below a certain threshold are set to zero. When the image is then transformed back into the original image format, the noise will have been reduced, whilst

the important information, such as the edges, will have been retained. To ensure that the right amount of de-noising is performed, whilst retaining useful image content, an iterative process is performed, and the final reconstructed image is the one which is outputted by the scanner.

Compressed sensing can be used to accelerate many different MRI scans, without impairing image quality. This was shown in a study by Delattre *et al*. comparing compressed sensing with conventional MRI sequences in 10 different applications, including brain imaging (3D FLAIR, 3D T_1-weighted with gadolinium), liver (3D T_1 mDixon with gadolinium) and MSK (3D PD VISTA SPAIR)[50]. For most applications, no statistically significant differences in image quality or artifacts were found between the compressed sensing and conventional sequence. Additionally diagnostic findings were similar, apart from four cases where motion had impacted the images (which occurred on either the conventional or compressed sensing sequence).

However, caution does need to be taken. Artifacts may occur, unique to compressed sensing. In a study of Compressed SENSE (Philips), three artifacts were identified as unique to Compressed SENSE scans[51]. There were titled "wax-layer" artifact, "streaky linear" artifact and "starry sky" artifact. However, the authors noted that the artifacts were only "slightly disturbing" to image quality. Also, they were not present on high contrast to noise sequences such as 3D T_2 driven equilibrium, 3D time of flight or 3D phase contrast angiography. They were, however, present on images from common clinical sequences such as 3D FLAIR and 3D T_1 m-Dixon TFE and were aggravated by patient movement. In a study of neuroimaging with compressed sensing, a "global ringing" artifact and image blurring was found to affect some cases, and they suggested that would hinder the diagnostic utility of the images[52]. They suggested that these particularly impacted cases with coarser resolution and/or higher acceleration factor. In the study by

Delattre *et al.* the elbow MRI with Compressed SENSE had a lower contrast to noise ratio than the traditional MRI[50]. However, the amount of denoising can be determined by the user, which could help with improving this.

Siemens and GE scanners have options for compressed sensing: on Siemens it is called Compressed Sensing and on GE HyperSense. Compressed sensing can be combined with traditional parallel imaging to further accelerate acquisition. On Philips scanners compressed sensing and parallel imaging methods are combined into one product Compressed SENSE (although parallel imaging alone without compressed sensing is still an option).

The amount by which Compressed Sensing can accelerate the acquisition depends on the sequence and application but is typically higher than that which is possible using parallel imaging. For example, in the previously mentioned study by Delattre *et al.* acceleration factors used varied from 2.8 used for the brain 3D T_1-weighted with gadolinium to 9 for a bile duct MRCP scan[50]. Since compressed sensing is a newer method compared to Partial Fourier and parallel imaging, the achievable acceleration is the subject of ongoing research.

3.6. Quantitative MRI

In current clinical use, MRI is essentially qualitative, with radiologists making judgements based on how bright a region is compared to another on an image. This can result in high inter and intra-radiologist variations due to the subjective nature of the assessment. Additionally, differences in imaging parameters such as TR and TE can result in differences in contrast which the radiologist may not be aware of. Quantitative MRI (qMRI) aims to generate an image whose voxels have a physical dimension and meaning and which, as a result, factor out the variations due to the subjective nature of the

assessment, and differences in image formation. Examples of qMRI parameters include the time constants T_1, T_2 and T_2*. To generate maps for these parameters a series of images with different T_1, T_2 and T_2* weightings can be acquired and fitted to the corresponding relaxation model. Each pixel will then have a specific T_1, T_2 or T_2* value, typically measured in ms. Sometimes it is clinically more relevant to give relaxation rates R_1, R_2 and R_2* estimations which are $1/T_1$, $1/T_2$ and $1/T_2$* respectively. R_1, R_2 and R_2* are typically measured in s^{-1}.

3.7. Iron and MRI

Given that this thesis is focused on detecting iron levels in the liver, this sub-section focuses on the impact of iron in the liver on MR imaging.

Ferritin and hemosiderin, which store iron, are superparamagnetic. This means that they are single-domain magnetic particles which, when placed in an external magnetic field, develop strong internal magnetization (but do not retain net magnetization once that field is removed). They therefore create magnetic field distortions to the external B_0 magnetic field. These field distortions impact both T_2 and T_2*. For conciseness I will refer to iron instead of ferritin and hemosiderin. T_2* is shortened by the presence of iron, directly due to the impact of iron on the B_0 field[53]. GRE sequences do not eliminate impacts of B_0 inhomogeneities, and hence are used for T_2* sequences, both weighted and quantitative. T_2 is also shortened by the presence of iron, despite the SE sequences used for T_2 imaging having a refocusing pulse to eliminate the impact of B_0 inhomogeneities. The mechanism for T_2 shortening is not well understood[53], and several explanations have been suggested. One suggestion is that when the protons undergo diffusion between the excitation and echo they each experience a different field (as the presence of iron creates B0 inhomogeneities). This results in incomplete rephasing after the refocusing pulse[53].

Another suggestion is that protons bound to iron-containing proteins undergo enhanced relaxation due to the iron electrons. This enhanced relaxation is then transferred between the bound proteins and bulk water through chemical exchange[53].

These properties mean that MRI can be used to detect and monitor iron overload in patients with diseases such as haemochromatosis and transfusion dependent thalassaemia[40,54]. The presence of high levels of iron in the liver, i.e., iron overload, can be assessed using T_2-weighted and T_2*-weighted imaging: manifested as reduced signal intensity. Shorter T_2* relaxation can also be detected using dual echo GRE sequences: in-phase and opposed-phase imaging (with echo times so water and fat are in and out of phase): if there has been a large decrease in signal between the two echoes then the patient is likely to have iron overload. However, these qualitative techniques cannot be used to reliably estimate the concentration of iron in the liver. qMRI provide methods whereby measurements made from a sequence can be directly converted into iron concentration. T_2 or R_2 mapping can be used to estimate LIC[55]. Multiple SEs are acquired with different TEs. The signal can be plotted against TE and an exponential curve fitted to give R_2. A calibration curve is available to then convert R_2 to LIC (at 1.5 T)[55]. However, the sequence for R_2 mapping requires a long acquisition time, of up to 20 minutes[56]. Not only does this have the disadvantage of resulting in longer scan times, it also can be prone to motion artifacts[56]. T_2* or R_2* mapping can also be used to estimate LIC. Multiple studies have produced calibration curves for converting 1.5 T R_2* into LIC[57–59].

Acquisitions for R_2* mapping are typically multi-echo (ME) GREs, which involve one excitation of the RF pulse, followed by multiple echoes, as shown in Figure 13. Multiple echoes are made by the rephasing and dephasing of the signal. The acquisition time of the ME GRE sequences used to obtain R_2* maps are much shorter than the acquisition time for R_2 mapping, and can usually be acquired in a breath-hold[56].

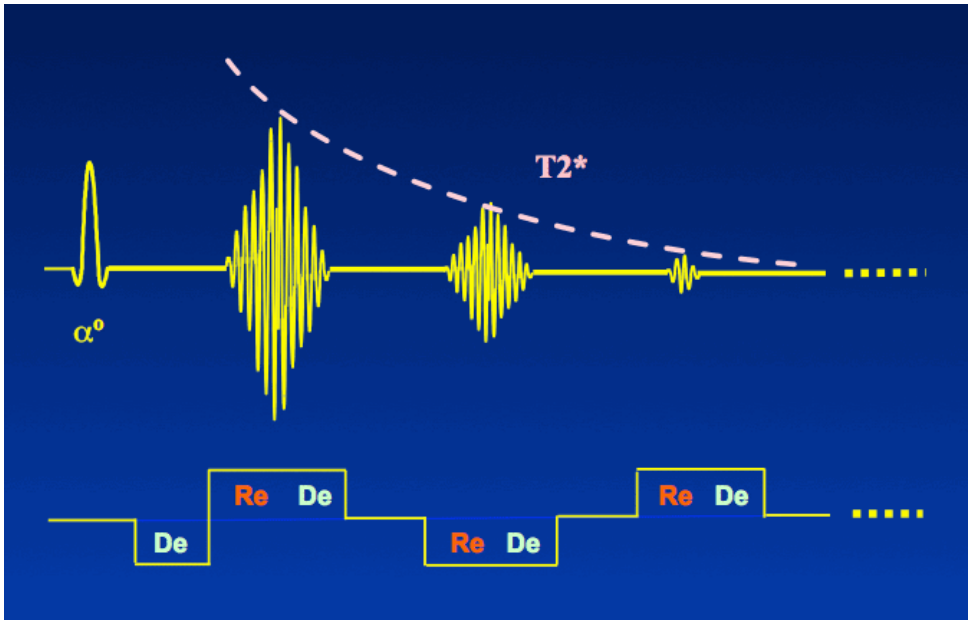


Figure 13 ME GRE: one RF pulse is applied, followed by dephasing and rephasing of the signal using gradients, to form multiple echoes. Courtesy of Allen D. Elster, MRIquestions.com

Clinically, 1.5 T and 3 T scanners are typically used. 3 T scanners can be seen as advantageous as the SNR is directly proportional to magnetic field strength. R_2* is also, in theory, directly proportional to field strength and this has been verified in several studies [60–62]. T_2* decay being shorter at 3 T can be both a limitation and an advantage. As iron shortens T_2*, a high field is good for detecting low iron as it can reduce acquisition times relative to low fields. However high fields can make it more difficult to detect iron overload as the compounded shortening of T_2* can mean that the signal has decayed away before the first echo. Increased B0 inhomogeneity, for example at air-tissue interfaces, is a disadvantage which impacts both high and low iron cases.

To estimate R_2*, an exponential decay curve can be fitted to the signal. However, in the liver , the signal is not a simple exponential decay: if there is fat present in the liver there will be oscillations as water and fat signals come in and out of phase. Therefore, in the liver, a more accurate estimation of R_2* can be made based on the following signal equation for each echo, which also includes the impact of B_0 field inhomogeneity:

$$s_i = \left(\rho_W + \rho_f \cdot \sum_p a_p e^{j2\pi f_p t_i}\right) \cdot e^{j(2\pi\psi t_i + \phi_0)} \cdot e^{-R_2^* t_i}$$

s_i is the signal for a particular echo, ρ_W and ρ_f are the amounts of water and fat. $\sum_p a_p e^{j2\pi f_p t_i}$ reflects the multi-peak fat spectrum model (a_p is the amplitude of the spectral peak p relative to the water peak and f_p is the difference in precession frequency of water and the fat peak p). $e^{j(2\pi\psi t_i + \phi_0)}$ represents the impact of field inhomogeneity (captured by parameter ψ) and the phase associated with the water and fat (considered equal and given as ϕ_0). R_2^* is the relaxation rate and n_i is the noise. ρ_W, ρ_f and R_2^* are unknown.

This equation can be used to estimate proton density fat fraction (PDFF) as well as R_2^*. PDFF is given by

$$PDFF = \frac{\rho_f}{\rho_W + \rho_f} \times 100$$

PDFF has many important applications in liver imaging as steatosis is an important feature in liver diseases[63]. Much of the work in developing robust estimation of R_2* is closely related to work related to developing robust estimation of PDFF. Originally PDFF was estimated using two echoes, in and out of phase, but since the publication of the "iterative decomposition of water and fat with echo asymmetric and least-squares estimation" (IDEAL) method in 2005[64] it has become increasingly common to use ME GRE data and model-based fitting to estimate PDFF (and R_2* as the same signal model includes R_2* too).

If this equation is used to estimate ρ_W, ρ_f and R_2^* then they rely on field map estimation. The IDEAL method is an example of a method for estimating PDFF and R_2* where the field map is estimated[64]. In IDEAL the field map, water, and fat are estimated using an iterative least-squares region-growing algorithm. Field map estimates from

nearby pixels are used to improve the initial guess of the field map. To estimate the field map, phase information is required. One problem with this is that the phase data is not always reliable[53]. Additionally, estimating the field map is a non-trivial optimisation problem, so there is potential for the wrong minima to be chosen[65]. This can result in fat-water swaps. Given that this method uses the magnitude and phase images, it is called a complex-based method. However, an alternative method is to use only the magnitude images i.e., a magnitude-based method. The equation above becomes

$$|s_i| = \left|\rho_W + \rho_f \cdot \sum_p a_p e^{j2\pi f_p t_i}\right| \cdot e^{-R_2^* t_i}$$

By using the magnitude signal model, the estimation of the field map is no longer required. Using the magnitude data is therefore desirable, but this has its own disadvantages. One of the main disadvantages is that using the magnitude data results in a non-Gaussian distribution of the noise. Therefore, at low levels of signal there is a "noise floor" with non-zero mean[53]. However, there are methods to mitigate this, such as using echo truncation: discarding of echoes below the noise floor[53]. Another disadvantage to consider is the loss in SNR due to the discarding of the phase images. Additionally, in most magnitude only techniques, it is not possible to quantify PDFF over 50%[63]. However, it was decided that the disadvantages of using the complex method (i.e., the requirement for phase images and the issues with field map estimation) outweighed its advantages, so throughout this work a magnitude only method was used (unless otherwise stated, for example if using inline T_2* processing on the scanner).

The specific method used was a technique called MAGnitude Only (MAGO). One reason for use of this method is that it forms part of the regulatory approved (FDA and CE) software LiverMultiScan of my industrial partner, Perspectum, making any outputs of this work easier to translate into the clinic. The advantage of MAGO above other

magnitude techniques is that, although not directly relevant to this work, it enables quantification of PDFF across 0-100%[65]. The disadvantages of using a magnitude method i.e., the noise floor and lower SNR, can be mitigated through echo truncation and boosting the signal (e.g., by using more echoes) respectively.

To estimate the unknowns, it applies an unconstrained nonlinear least-squares fitting model twice, starting with different initial values of ρ_W and ρ_f based on 0% and 100% proton density fat fraction (PDFF). The initial value for R_2* is 50 s^{-1} (T_2* of 20 ms). This results in two solutions, labelled Solution 1 and Solution 2. Pixels are then selected from either Solution 1 or 2 based on the residual sum of squares following regularisation steps to ensure connectivity and 'realistic' T_2* values (between 0 and 200 ms). This yields T_2* (and hence R_2*) and PDFF maps. Additionally, echo truncation is applied so that later echoes are excluded if the signal is below a specified threshold. A flow chart of the MAGO process is shown in Figure 14.

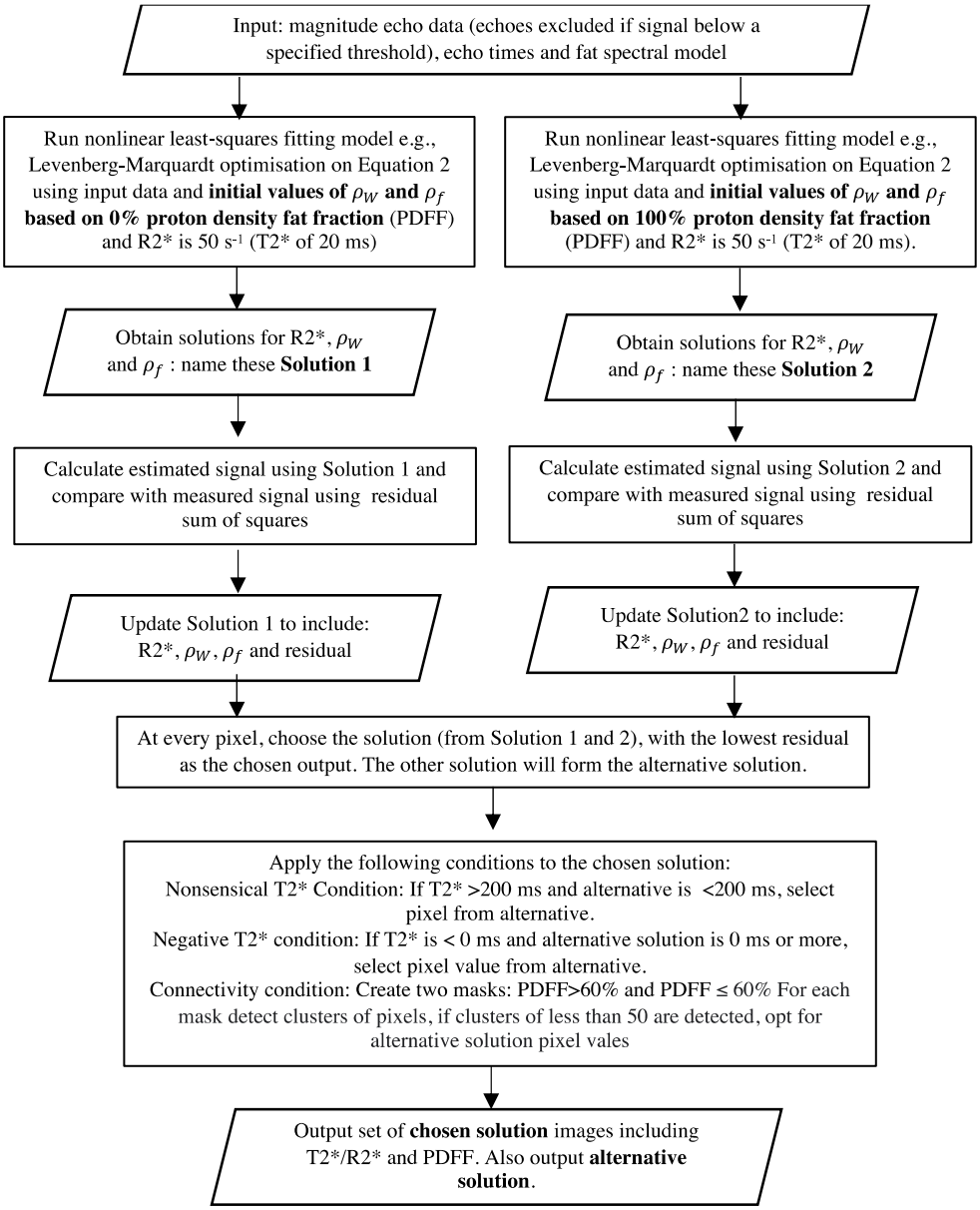


Figure 14 Flow chart of the MAGO process

A 2D ME GRE sequence used several times in this thesis is called Magnitude-Only thin Slice T_2* (MOST) developed by Perspectum. Corresponding parameters are shown in Table 3. The acquisition can be obtained in a breath-hold of approximately 10 s. Parameters are different between field strengths and the parameters also vary slightly between scanners.

Parameter	Siemens 1.5T	Siemens 3T	Philips 1.5T	Philips 3T	GE 1.5T	GE 3T
Field of View (mm x mm)	440 x 357		440 x 378		440 x 352	
Acquisition Matrix	128x104		142x140		128x128	
Acquired Resolution (mm x mm)	3.44 x 3.43		3.10 x 2.7		3.44 x 2.75	
Reconstructed Resolution (mm x mm)	1.72 x 1.72					
Acquired Slice Thickness (mm)	3.00					
Reconstructed Slice Thickness (mm)	3.00					
Number of Slices	1		1		7 (0 mm slice gap)	12 (0 mm slice gap)
Number of Repeats	7	12	7	12	1	
TR (ms)	~21	~11	~21	~11	~21	~11
Number of Echoes	8					
Initial TE (ms)	2.38	1.23-1.33	2.37	1.18-1.32	1.82 –2.22	1.06 –1.54
Final TE (ms)	19.04	9.84	18.96-19.04	9.44-9.65	18.5-21.75	9.67-10.15
Echo spacing (ms)	2.38	1.23	2.37-2.38	1.18-1.19	2.38-2.79	1.23
Flip Angle (degrees)	15	9	15	9	15	9
Parallel imaging	GRAPPA (2-fold with 35 reference lines)		SENSE (2-fold)		ARC 2.0	
Acquisition Duration (s)	~10					

Table 3 Sequence parameters, range/approximation due to slight variation between scanners.

The sequence has been developed to have high SNR and to be robust against artifacts. For example, at 1.5 T and 3 T the sequence involves acquiring respectively 7 and 12 repeated slices (at the same location). Having multiple repeats improves the signal to noise ratio (SNR). Additionally, repetitions where there is movement can be discarded. The algorithm used for processing the raw echoes first involves the generation of T_2* maps for each repeat, then the automatic selection of a subset of the most similar repeats (subset sizes are 4 for 1.5 T and 7 for 3 T). This is followed by averaging of the raw echoes of that subset of repeats, and the generation of a T_2* map from those averages. Echo truncation is performed if signal is below a certain threshold. The 3 mm slice thickness is beneficial as susceptibility differences at air/tissue interfaces (e.g., between lung and liver) can result in T2* shortening in the liver, and a thin slice thickness reduces the impact of this.

3.8. MRI and Detection of Iron Sparing

3.8.1. MRI for detection of iron sparing

To date, the primary clinical use of MRI to determine the presence of iron in the liver has been for diagnosing and monitoring iron overload[40,54]. However, LI-RADS has iron sparing in a solid mass as an "ancillary feature favouring malignancy in general"[6]. As an ancillary feature it can be used for aiding detection and increasing confidence of diagnosis of HCC. It can be also used to upgrade the LI-RADS category, with the exception of upgrading LR-4 (Probably HCC) to LR-5 (Definite HCC). It is limited to cases which have iron overload. Iron overload is typically defined to be LIC of 3.2 mg/g dw or above[56], which, as shown from Ebara's two studies, is not typical of patients with HCC, where the liver parenchyma mean iron concentrations were 1.42 mg/g dw[20] and 1.66 mg/g dw[17]. The LI-RADS MRI protocol only includes qualitative imaging and not quantitative imaging which may limit the scope of this ancillary feature further. The LI-RADS manual does say that R_2* mapping can be used to determine iron sparing (lower R_2* in HCC compared to liver tissue), but data for R_2* mapping is not required by LI-RADS[6].

There is limited literature on the ability of MR imaging to detect iron sparing. Ideally there would be a biopsy paired study of MRI imaging with iron content of tumours and livers measured. The closest study we have to this is by Terada *et al.*[66]. The study was not directly related to iron sparing in HCC but did look at differences in iron between nodules and surrounding tissue, primarily at adenomatous hyperplastic nodules (AHN) (not cancerous) and surrounding regenerative nodules. Some HCCs were included but iron was not quantified in those cases. This was an in vitro study, with samples scanned at 1.5 T to give T_1-weighted, T_2-weighted and proton density images. Apparently, all images were almost the same[66]. Differences in intensity between nodules were reported. The

results for cases where iron concentration was measured in AHN and regenerative nodules is given in Table 4.

Case	Iron (mg/g dw) AHN	Iron (mg/g dw) Regenerative nodules	Difference in intensity between AHN and Regenerative nodule
1	5.00	2.80	Yes, AHN hypointense
2	2.80	2.10	Yes, AHN hypointense
3	1.00	0.77	No
4	1.90	1.60	No
5	0.29	0.26	No
6	1.40	1.60	No
7	1.30	1.20	No

Table 4 Iron Concentration in adenomatous hyperplastic nodules and regenerative nodules, and whether there was a difference in intensity on MRI imaging, results from Terada *et al.*'s study[66]

This study indicates that large differences in iron (e.g., 5.00 mg/g dw vs 2.80 mg/g dw) can be seen using weighted imaging on MRI but smaller differences (e.g., 1.90 mg/g dw vs 1.60 mg/g dw) cannot be seen. It should be noted that a limitation of this study is that it was *in vitro* and not *in vivo*.

Although not biopsy paired, there have been two studies investigating whether HCCs can be detected on T_2*-weighted imaging, given that the HCC may have iron sparing. Both these studies involved ME GRE sequences, with radiologists having the echo images (i.e., the T_2* weighted images) to identify the tumour. They also had access to T_2-weighted imaging to determine if the lesion was cystic (i.e., benign) or solid. T_2* maps were also made from the ME imaging, but these were not used for detecting the HCC. In the first study, 14 patients with 25 HCCs were imaged with a ME GRE with 6 echoes[67]. Confirmation of HCC was made using gadolinium enhanced images. Sensitivity for detecting HCC using T_2* weighted images was 60%[67]. T_2* was significantly higher in the liver parenchyma in patients where the HCC was not detected, 27.8±2.2 vs. 21.9±2.1 ms (at 1.5 T)[67]. This corresponds to lower iron concentrations, indicating that it was more difficult to detect HCC on the T_2* weighted images in patients with lower iron in the parenchyma.

The second study using T_2* weighted imaging included 73 patients with 98 HCCs[68]. The ME GRE had 3 echoes, 9.5 ms, 19.3 ms and 29 ms. In this study the radiologist identified HCC, firstly using 9.5 ms echo time, then using 19.3 ms echo time a week later, then 29 ms a week after that. HCC confirmed on pathology. 37 of the 98 HCCs were identified at TE = 9.5 ms, 45 HCCs at TE = 19.3 ms and 41 HCC at TE = 29.0 ms[68]. Overall, 57 (58%) were identified at least one TE. No HCCs were visible on T_2*-weighted imaging for patients with hepatic iron content <1.0 mg/g dw measured by T_2* (Wood's conversion) (33 HCCs), whereas for hepatic iron content greater than 1.0 mg/g dw 88% of HCCs (57 of 65 HCCs) were visible[68].

The results from the T_2* weighted imaging studies seem to indicate that HCC tumours are harder to detect in lower iron cases. The reason for this firstly could be because iron differences are smaller in these cases. Iron differences are likely to be smaller as there is less iron initially, and this is shown in the plot from one of Honda *et al.*'s studies[18], shown in Figure 6, where the biggest differences in iron between the HCC and the liver is shown in the cases with the largest iron overload. Not being able to detect small iron differences is consistent with the AHN/regenerative nodule study, where smaller iron differences could not be detected[66]. As well as the challenge of detecting smaller iron differences, the slow T_2* decay at lower iron may also contribute to the contrast between HCC and iron not being enough at the echo times provided to enable HCC detection. The impact of choice of different echo times is shown by the second study referred to above, where different numbers of HCCs were detected dependent on the echo time[68]. An additional limitation of the sequences used in the T_2*-weighted studies was that 15 slices with 10 mm slice thickness were used, with a slice gap of 2 mm[67,68]. The thick slices mean that through-plane T_2* shortening is likely near the lungs, due to susceptibility. Additionally, the thick slices could make detecting smaller tumours (1-2 cm) challenging.

Using qMRI and optimising acquisition sequences using may enable the detection of smaller differences in iron and enable better identification of iron sparing, if present. qMRI has the advantage that is does not depend on visual evaluation, and so smaller differences can be detected. Additionally, for weighted imaging, the optimal echo times depend on the difference in iron to be detected (as shown in the previously mentioned study), but with qMRI a range of echoes are used to fit the curve, making it more suitable for different amounts of iron. Optimisation could include improving signal to noise ratio (SNR), resolution and choosing specific echo times for this application, all of which could improve the detection of iron sparing.

3.8.2. Proposal

The available evidence seems to suggest that iron sparing does occur in HCC, but further verification of this is required. Additionally, if iron sparing is a feature of HCC which occurs soon after malignancy begins, then a non-invasive tool to detect this could be used, alongside other techniques, to help with early diagnosis of HCC. A qMRI tool could therefore be used to meet both needs: firstly, to verify if iron sparing happens, and if so, secondly, to be used as part of HCC detection. Therefore, I proposed to develop an MRI sequence and analysis protocol which would enable iron sparing detection, and this is the focus on this project and the rest of this thesis.

For qMRI techniques, there is a choice between using R_2 or R_2* mapping. Given it takes up to 20 minutes[56] to acquire the data for R_2 mapping, I decided to use sequences for R_2* mapping in this project.

3.8.3. Considerations when using R_2* mapping for iron sparing

If applying R_2* mapping to detect iron sparing, of primary importance is being able to detect R_2* differences corresponding to the differences in iron we expect to see between the HCC and liver. Although R_2* or T_2* can be used (as $R_2^* = 1/T_2^*$ in the context of

evaluating differences in LIC it can be more helpful use R_2*, due to the linear relationship of R_2* and LIC, unlike T_2*, see Figure 15 (calibration curves from Wood *et al.*[57]).

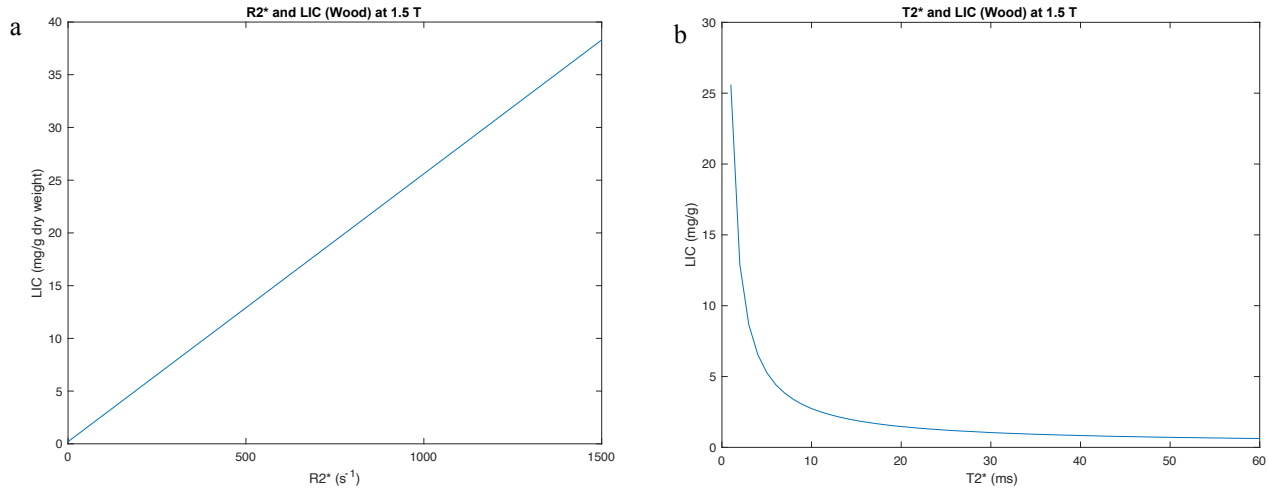


Figure 15 a) R_2*-LIC calibration curve, b) T_2*-LIC calibration curve, at 1.5 T, based on Wood's study

As can be seen from the graphs, if a certain ΔLIC is selected, this will correspond to the same ΔR_2*, regardless of whether two higher values or two lower values of LIC are selected. However, due to T_2* being the reciprocal of R_2* this is not the case for T_2*. The same difference in LIC will correspond to much larger difference in T_2* at lower values of LIC, compared to higher values. This difference in the relationship with change in LIC between R_2* and T_2* can be shown algebraically. Using Wood's equation, for R_2*:

$$LIC = 0.0254(R_2{}^*) + 0.202$$

$$\frac{dLIC}{dR_2{}^*} = 0.0254$$

And hence change in iron is independent of R_2*. For T_2*:

$$LIC = 0.0254\left(\frac{1}{T_2{}^*}\right) + 0.202$$

$$\frac{dLIC}{dT_2{}^*} = -0.0254\left(\frac{1}{T_2{}^{*2}}\right)$$

Hence, we can see that change in LIC is not independent of T_2*.

To give an idea of the impact of the different relationships of R_2* and T_2* with change in LIC, consider the example of a difference of 0.1 mg/g dw, firstly increasing from 3 mg/g dw to 3.1 mg/g dw. Using Wood's equation, the corresponding difference in R_2* is 3.94 s^{-1}. The corresponding difference in T_2* is -0.313 ms. Then taking the difference between 1 mg/g dw and 1.1 mg/g dw, the difference in R_2* is also 3.94 s^{-1}. With T_2*, the difference is -3.54 ms. Hence, it could be misleading to use T_2* when assessing differences in LIC (both within and between patients): the same LIC difference will correspond to larger differences in T_2* at lower LIC levels. An example of the impact of this in two healthy volunteers is shown in Figure 16 and Table 5 . Two ROIs were placed, depicted as white and green circles in the figure, on T_2* maps and the difference between them in terms of T_2*, R_2* and LIC was calculated. Although the difference in T_2* between the two ROIs is much greater in the volunteer with longer T_2* (057) compared to the volunteer with shorter T_2* (125) the difference in R_2* and LIC is much less.

In this thesis both T_2* and R_2* are used, but R_2* is the measurement of choice.

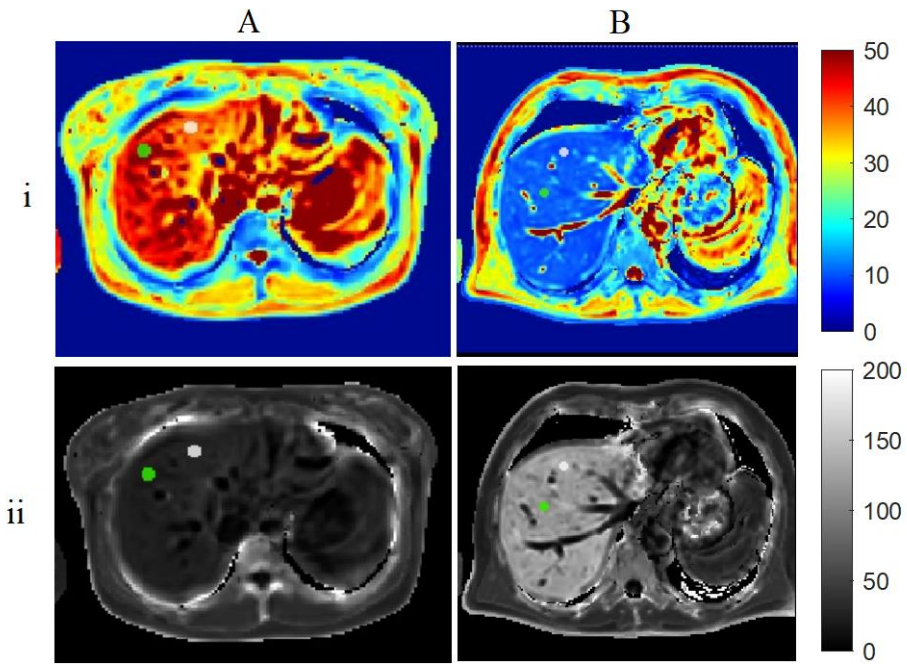


Figure 16 (i) T_2* and (ii) R_2* maps from two volunteers (labelled A and B) scanned on a Siemens 1.5 T scanner. ROIs are shown in white and green.

Volunteer	Label name	ROI Mean T_2* (ms)	ROI Mean R_2* (s^{-1})	ROI Mean LIC (mg/g dw, Wood's equation)
A	White ROI	39.33	25.44	0.85
	Green ROI	49.83	20.12	0.71
	Absolute difference between white and green ROIs	10.49	5.32	0.14
B	White ROI	7.83	127.82	3.45
	Green ROI	9.60	104.30	2.85
	Absolute difference between white and green ROIs	1.77	23.52	0.60

Table 5 ROI values for the volunteers, with R_2* converted to LIC using Wood's calibration curve

Another consideration is whether the values should be kept as R_2* or converted to iron concentration when assessing for iron sparing. Given that the tissue properties in cirrhotic livers and HCCs are different to that of the liver tissue from which calibration curves were produced (typically patients with diseases such as thalassemia[57,58]), it is true that the conversion may not apply. Other factors may influence the T_2* decay such as deoxygenation[69] and the presence of fat[70]. It is true that some studies have suggested that R_2* in tumours could be linked to deoxygenation, including two comparing R_2* in HCC compared to benign lesions[71,72], which found increased R_2* in HCC compared the benign lesions. However, comparison was not made between HCC and liver tissue within the

same patient: in that scenario changes in iron may dominate over changes in oxygenation, especially when we consider that ferritin and hemosiderin are superparamagnetic, whereas deoxyhaemoglobin is paramagnetic. Additionally, a recent narrative review[73] stated that so far the results are inconclusive as to whether there is severe hypoxia in HCCs and stated that "The widespread remark "hypoxia is a typical feature of HCC" is not supported by reliable, published data and should not be used.[73]" It may be that deoxygenation does impact the signal to some extent, but to develop the tool and assess it, it is useful to use iron conversions, so in this project I assumed that the iron conversion would apply.

3.8.4. Requirements for detection of iron sparing

To enable assessment of existing sequences, guide the optimisation process and allow for evaluation of the optimised sequences a clear set of requirements for a sequence which could be used for detecting iron sparing is necessary. These need to be based on the available literature and have a clear rationale behind each one. I therefore drew up the following requirements:

1) Measure iron concentration from 0 mg/g dw to 3 mg/g dw +/- 0.1 mg/g dw
2) detect differences of 0.2 mg/g dw or more of iron concentration within the liver (for tumour identification)
3) Detect tumours of size 1 cm or more.
4) Scan breath-hold length of 16 s or less, maximum of 3 breath-holds
5) Full coverage of the liver
6) Deployable on standard clinical MRI scanners without modifying vendor software

Although the key objective is to be able to detect iron differences, to do this is it necessary to ensure that the sequence can enable measurement of iron in the range we are expecting. If not, then it is unlikely we will be able to detect iron differences. I therefore specified that the tool should be able to measure iron concentration from 0 mg/g dw to 3

mg/g dw +/- 0.1 mg/g dw (Requirement 1). This is based on studies where dry weight liver iron concentration was measured in HCC and liver parenchyma. For example in one of the studies by Ebara *et al.*[17] mean +/- standard deviation (SD) of iron concentration in HCCs was 1.3 +/- 0.1 mg/g dw vs 1.7 +/- 0.1 mg/g dw in liver tissue[17]. In the study by Tashiro *et al.*[19] geometric mean was 0.4 mg/g dw in HCC with geometric SD of 2.5 and geometric mean was 0.6 mg/g dw with geometric SD of 3.1[19] in the liver. Dry weight has been used as that is what the calibration curves for R_2* conversion to iron generally use[57,58]. Wet weight studies were not used to inform this requirement as for this conversion between wet weight and dry weight would be necessary, and it is unclear which conversion should be used as different conversion factors between wet weight and dry weight have been proposed[74]. The decision of allowing for +/- 0.1 mg/g dw is based on *Requirement 2* (being able to detect 0.2 mg/g dw differences in the liver): if there are more than 0.1 mg/g dw errors in different locations in the liver this will make detecting 0.2 mg/g dw differences difficult. Additionally, global differences (i.e., a bias) of more than 0.1 mg/g dw from true iron concentration is not desirable, because the tool may be useful for monitoring iron content in the liver or a nodule over time and not just differences throughout the liver, for example to monitor iron content in a dysplastic nodule. If there was a bias on different scanners, then comparisons could not be made if a patient is monitored over time on different scanners, or patients for the same study are scanned on different scanners. As we are looking at small changes in concentration in the case of iron sparing, then +/- 0.1 mg/g dw seems a reasonable threshold.

There is limited literature on the difference between liver parenchyma and tumour concentrations, so deciding on a requirement for the measurable difference was difficult. In one of the studies by Ebara *et al.* [17], the mean iron concentration was 1.3 +/- 0.1 mg/g dw in the HCC, compared to 1.7 +/- 0.1 mg/g dw in the liver. In individual patients there

will be a huge range of variation in iron differences, as shown in the figure from Honda *et al.*'s study, Figure 6. This study used wet weight iron concentrations, not dry weight, and as mentioned previously, I avoided converting between wet weight and dry weight due to the range of conversion factors available. However, using Ebara's study and seeing that some of the differences between tumour and liver were very small (Figure 6), I decided 0.2 mg/g dw differences would be a suitable threshold, and so this is Requirement 2. This means that ideally errors of less than 0.1 mg/g dw differences would be seen in the liver (as specified in Requirement 1).

An application of this tool is for use in early diagnosis of HCC. Tumours are classified as very early stage HCC when less than 2 cm [9], so ideally early diagnosis tools would detect tumours 0-2 cm. However, a lower limit of 1 cm was made based on pragmatism and the knowledge that the detection of tiny, sub-cm lesions will be impossible with the current technology. Therefore Requirement 3 is to detect tumours of size 1 cm or more.

HCC patients are often elderly and/or in poor health, and therefore may only be able to manage short breath-holds. Additionally, time in the scanner needs to be minimised, both for patient comfort and cost purposes. Hence Requirement 4 is to have a scan breath-hold length of 16 s or less, with a maximum of 3 breath-holds. An alternative would be to have a respiratory triggered sequence to avoid the requirement of breath-holding, but these take longer and could result in large motion artifacts if the patient moves (additionally to breathing) during the acquisition.

Tumours may occur anywhere within the liver, hence Requirement 5: full coverage of the liver. It should also be noted that good quality coverage is required.

Novel pulse sequences may provide enhanced sensitivity; however, this would limit the utility of this sequence as implementation across different scanners at scale would be difficult. Therefore Requirement 6 is that the sequence should be deployable on standard clinical MRI scanners without modifying vendor software.

3.9. Summary

MRI can be used to estimate iron and so I hypothesise that it could accurately measure iron sparing in HCC. However, using MRI for such a purpose will require further development to meet the requirements for this application. These requirements were outlined at the end of this chapter.

4. Iron conversion methods: An evaluation

4.1. Introduction

In Chapter 3 I specified iron concentrations and differences in iron concentrations which are necessary for detecting iron sparing, leading to the conjecture that R_2* can be used to detect these concentrations. A key component of developing and assessing a sequence for detecting iron sparing is conversion from R_2* to iron concentration. Multiple studies have produced calibration curves for the conversion of R_2* measured at 1.5 T to LIC[57–59,75]. The validity of such conversions has been supported by the majority of the calibration curves being in close agreement.[59] Confidence in the precision and accuracy of 1.5 T-R_2* LIC is sufficiently high that it is used in clinical practice, for example for monitoring LIC in patients with thalassemia[40]. There are some, but currently fewer, methods available to convert R_2* data acquired with 3 T scanners to LIC (either directly[76], or via a 3 T R_2* to 1.5 T R_2* conversion[60]). Due to the limited number of studies at 3 T, the confidence in these methods is lower compared to the 1.5 T-R_2* LIC conversion.

To assess the utility of the pulse sequences developed for this thesis, I needed to be sure that the 3 T R_2*-LIC conversion gives an accurate measurement of iron. One way of verifying its accuracy is to compare with the established 1.5 T R_2*-LIC (which we assume is accurate, given that is has been assessed to have reliable precision and accuracy and is recommended for use in clinical practice[40]). Additionally, given that there is potential for patients to be scanned across different field strengths over time, and different field strengths to be used in the same clinical trial when detecting iron sparing, it is important that the results are consistent between 1.5 T and 3 T.

Although this work was developed for the specific application of detection of iron sparing, it also has much wider contributions to the clinical community and I also cover

these applications in this chapter. For example, being able to monitor iron concentrations in patients with diseases such as haemochromatosis and thalassemia at 3 T with the knowledge the LIC calculated will be the same as that at 1.5 T will give confidence to clinicians to use 3 T scanners and so open up a wider range of scanners for this purpose. It will also mean that 1.5 T and 3 T scanners can be used for scanning the same patient over time, for example for monitoring chelation therapy, and used alongside each other in clinical trials in these disease areas.

In the work reported in this chapter, 30 participants were scanned across 1.5 T and 3 T scanners, and LIC calculated from R_2* compared between the two field strengths. This data was initially acquired and analysed for a different purpose to this work (and the data collection was not performed by myself), but once I was aware of this data, I realised its potential use in this area. I therefore came up with an analysis plan of the data. I made all the major decisions for this project, including which statistical analysis to perform. A graduate intern at Perspectum helped with plotting of the graphs, and I guided him through this. A statistician from Perspectum was also consulted at various points to check what I was proposing was a valid approach. I evaluated the results and drew the conclusions.

4.2. Methods

4.2.1. Patients

30 participants were recruited as part of the "Repeatability and Reproducibility of Multiparametric MRI" study (Clinicaltrial.gov NCT03743272). The inclusion criteria were male and female subjects aged between 18 - 75 years, and ability to understand and sign a written informed consent form (ICF). A subset had previously been diagnosed with high liver fat or high iron. The trial had ethical approval (South-Central Oxford Research

Ethics Committee C (Ref: 17/SC/0205)) and all participants gave written informed consent. Participant characteristics are given in Table 6.

	Median (Range)	N
Gender		
Male		15
Female		15
Age (years)	36 (19-73)	
BMI (kg/m^2)	25.4 (19.5-35.3)	
Normal/Healthy Weight (18.5-24.9 kg/m^2)		14
Overweight (25.0-29.9 kg/m^2)		10
Obese (>30.0 kg/m^2)		6
Weight (kg)	79.2 (47-102.3)	
Height (cm)	173 (147-192)	
Reported healthy		19
Reported liver disease		11
Disease details (if available)		
Fatty liver		8
Unhealthy (AIH, PSC, Wilson's Disease)		3

Table 6 Participant characteristics

4.2.2. MRI protocol and Post Processing

Ten participants were scanned on one 1.5 T scanner and one 3 T scanner, and twenty participants were scanned on one 1.5 T scanner and on two 3 T scanners. The resulting number of data points (1.5 T and 3 T paired data) was 50. All participants had fasted for 3-5 hours before each scan.

Three major scanner vendors were represented (Siemens, Philips, and General Electric (GE)) and a variety of models were used. The scanners are shown in Table 7, with the number of participants scanned on each. It should be noted that "3 T scanners" are not always exactly 3.00 T: Siemens 3 T scanners are actually 2.89 T, while Philips and GE 3 T scanners really are 3.0 T.

	Field Strength	N
Siemens		
Prisma	3 T	20
Prisma Fit	3 T	10
Avanto Fit	1.5 T	10
Philips Medical Systems		
Achieva dStream	3 T	10
Ingenia	1.5 T	10
GE Medical Systems		
Discovery MR750	3 T	10
Optima MR450w	1.5 T	10

Table 7 Scanner vendor, model, field strength and number of scans per scanner

On each scanner the MOST sequence was used: the parameters are given in Section 3.7. The MAGO method was used to generate T_2*[65] maps, as described in Section 3.7.

Once T_2* maps were obtained trained analysts placed three circular ROIs, each 15 mm in diameter, in representative areas of the liver parenchyma, so as to avoid artifacts, large blood vessels and bile ducts. A median T_2* of the pixels included in the three ROIs was calculated. The median T_2* (units: ms) was then converted to R_2* in s^{-1}. An example of the T_2* maps for both 1.5 T and 3 T acquisitions are shown in Figure 17 with ROIs placed on the liver parenchyma.

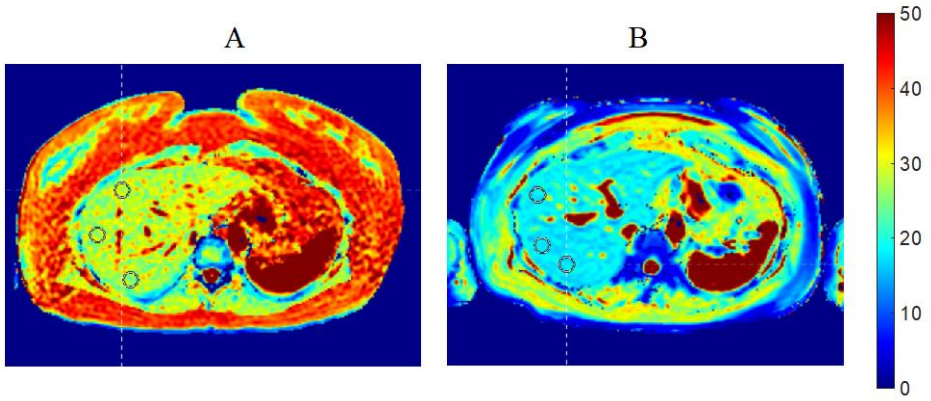


Figure 17 T_2* maps with 3 circular ROIs placed in liver parenchyma for the same patient at 1.5 T (left) and 3 T (right). Colour bar units: ms

4.2.3. 1.5 T LIC Conversion

Although the primary focus was to explore different 3 T-R_2* conversion to LIC techniques, several different 1.5 T calibration curves have been published. A previous study showed good agreement between a number of calibration curves[59], suggesting that

it may be acceptable to use any one of them. Both the Wood [57] and Hankins [58] equations for 1.5 T LIC were used to compare with 3 T LIC.

Wood's equation is:

$$LIC(mg/g) = 0.025\ R2^{*}_{1.5\,T} + 0.202$$

Hankins' equation is:

$$LIC(mg/g) = 0.028\ R2^{*}_{1.5\,T} - 0.455$$

The population in Hankins' study had LIC levels spanning from 0.6 mg/g dw to 27.6 mg/g dw[58], compared to Wood's which spanned from 1.3 mg/g dw to 32.9 mg/g dw[57]. Given that HCC patients typically have lower levels of iron[20] Hankins' study contains more patients in the iron range of relevance for detecting iron sparing, while Wood's population is more similar to a population with iron overload.

4.2.4. 3 T LIC Conversion

The methods compared for 3 T LIC conversion were

1. Direct conversion of 3 T-R_2* to LIC

For this the d'Assignies' calibration curve was used[76]. In their paper, d'Assignies' *et al.* give two equations, one of which was for patients with LIC of < 7.26 mg/g dw. We chose this equation for our study as it was the more relevant to the LIC range for the population we examined. The d'Assignies equation is given in their paper in μmol/g dw, so this was converted to mg/g dw using the atomic weight of iron. The d'Assignies equation becomes:

$$LIC(mg/g) = 0.0175\ R2^{*}_{3\,T} - 0.536$$

2. Two step conversion: firstly, conversion of 3 T-R_2* to 1.5 T-R_2*, then conversion of 1.5 T R_2* to LIC

We used the relationship found by Alam[60] to convert 3 T-R_2* to 1.5 T-R_2*. Alam's equation is:

$$R2^*_{3\,T} = 2.051\, R2^*_{1.5\,T} - 25.4276$$

Therefore, to convert 3 T R_2* to 1.5 T R_2* the equation is:

$$R2^*_{1.5\,T} = \frac{R2^*_{3\,T} + 25.4276}{2.051}$$

Furthermore, 3 T-R_2* to 1.5 T-R_2* conversion was also derived using the data from this study. For the conversion generated from the data in this study, an overall conversion equation was found, but also one conversion for "true 3 T scanners" i.e., Philips and GE, and one for 2.89 T scanners i.e., Siemens "3 T" scanners. Both Alam's conversion and the ones derived in this study were compared with 1.5 T-R_2* LIC.

Additionally, for method 2, both Wood's and Hankins' conversions to LIC could be used once 3 T R_2* is converted to 1.5 T R_2*, so both methods were used. The conversion methods are summarised in Table 8 and Figure 18.

1.5 T LIC conversion method	3 T LIC Conversion method
Wood	d'Assignies Convert 3 T to 1.5 T R_2* using Alam R_2* conversion, and convert to LIC using Wood Convert 3 T to 1.5 T R_2* using this study's R_2* conversion (pooled and vendor dependent), and convert to LIC using Wood
Hankins	d'Assignies Convert 3 T to 1.5 T R_2* using Alam R_2* conversion, and convert to LIC using Hankins Convert 3 T to 1.5 T R_2* using this study's R_2* conversion (pooled and vendor dependent), and convert to LIC using Hankins

Table 8 Summary of LIC conversion methods

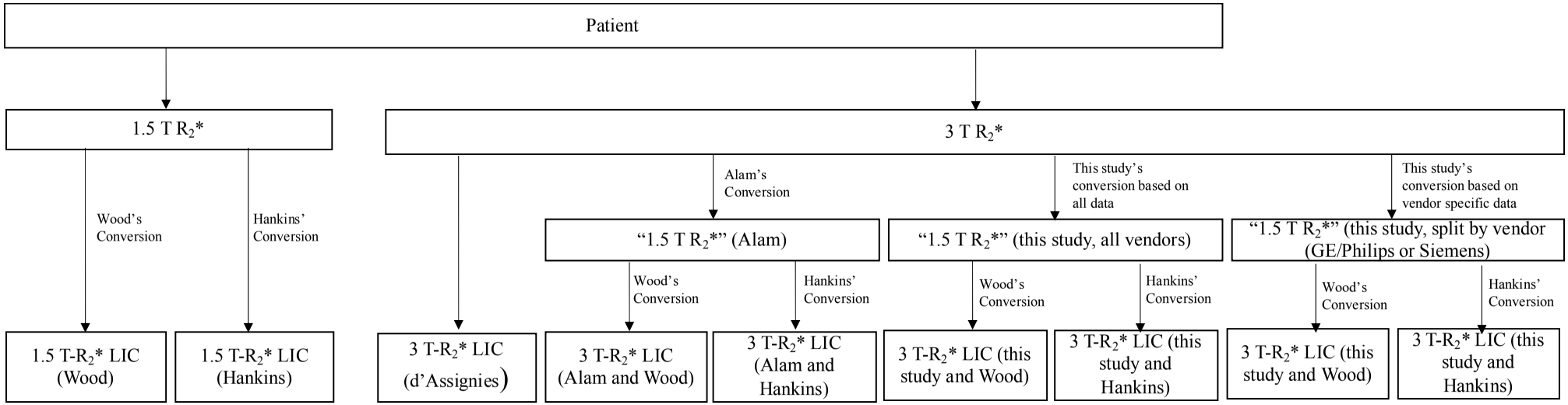


Figure 18 Diagram of the different methods to obtain LIC – the favoured approach in this thesis was 1.5 T-R2* LIC (Hankins) and 3 T-R2* LIC (Alam and Hankins)

4.2.5. Statistical Analysis

The relationship and correlation between the 3T R_2*-LIC and 1.5T-R_2* LIC metrics was assessed using linear regression and Spearman's rank-order correlation coefficients (rho). Agreement between the two measurements was assessed using the Intraclass Correlation Coefficient (ICC) and Bland Altman analysis (by calculating the bias and 95% limits of agreement (LoA)). The relationship and agreement were also assessed visually using scatter plots and Bland Altman plots with different colours and shapes chosen to distinguish between scanner types. For statistical tests, p-value < 0.05 was concluded to be significant.

4.2.6. Relevant thresholds for agreement

The level of agreement required between 1.5 T LIC and 3 T LIC depends on the application. For categorising iron overload, the categories are quite broad (e.g. 3.2 – 7.0 mg/g dw is mild overload, 7.0 – 15.0 mg/g dw is moderate overload and >15.0 mg/g dw is severe)[56]. Therefore, for diagnosis, a moderate amount of variability in LIC across field strengths is acceptable, e.g., ±0.5 mg/g dw.

However, for monitoring treatment in the same patient across different scanners at both 1.5 T and 3 T, a closer agreement between 1.5 T and 3 T LIC may be required. For example, in a study of patients receiving the iron chelator Deferasirox, the change in least squares mean LIC (derived using an Analysis of covariance (ANCOVA) model with treatment as factor and baseline LIC as covariate) was 1.95 mg/g dw over 52 weeks for patients receiving 5 mg/kg per day of Deferasirox, with standard error of the mean of 0.50 mg/g dw[77]. This works out at 0.45 mg/g dw every 12 weeks, so if monitoring of LIC was done on a 12-weekly basis then a variability of 0.23 mg/g dw in results across field strengths may be acceptable.

For detection of iron sparing, it is also desirable to have a close level of agreement, especially as we are interested in low levels of iron concentration and small iron concentration differences. As stated in Section 3.8.4, therefore the requirement is to measure iron concentration with +/- 0.1 mg/g dw accuracy. Ideally, we would have this level of agreement between 1.5 T-R_2* LIC (which we assume to be accurate) and 3 T-R_2* LIC (which we are assessing).

4.3. Results

The relationship between R_2* at 3 T and 1.5 T, with fitted linear regression lines and equations, alongside Alam's conversion, is shown in Figure 19. The linear regressions found from the data in this study are similar to Alam's, but both the slope and intercept are slightly different. The equations for the linear regressions were used in one of the methods for obtaining 3 T-R_2* LIC.

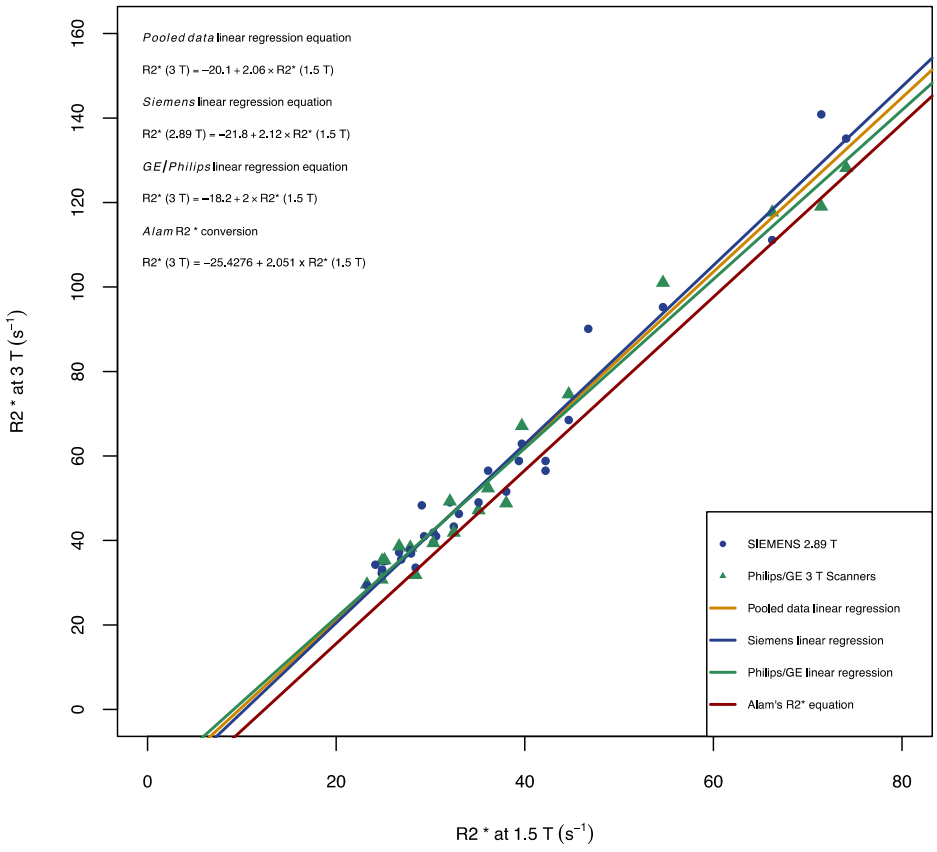


Figure 19 3 T R_2* plotted against 1.5 T R_2* with Alam's conversion line, and the linear regression lines using the data.

Figure 20 and Figure 21 show examples of scatter plots and Bland Altman plots comparing LIC from 1.5 T (Wood and Hankins) with LIC from 3 T (d'Assignies and two-step method using Alam and Wood/Hankins). Further plots comparing the other methods are given in Appendix A. Spearman's rho and ICC values are shown in Table 9 and Table 10. Bias and LoA are given in Table 11.

The scatter plots and Table 9 show that 3 T-R_2* LIC and 1.5 T-R_2* LIC had strong correlation, regardless of the method used. However, the regression lines for d'Assignies' 3 T-R_2* LIC are quite different to the identity line, unlike the other 3 T-R_2* LIC methods, where the regression lines are close to the identity line. Table 10 (which gives the ICCs) confirms that d'Assignies' 3 T- R_2* LIC had a much lower agreement (and not always significant) with 1.5 T-R_2* LIC compared to the other methods. For the other methods, the agreement given by ICC was always excellent, as well as being significant, and the regression lines are close to the identity line. The method involving the 3 T to 1.5 T R_2* conversion equations found in this study (both vendor specific and pooled) gave the highest ICCs. However, the difference in ICCs when using this study's conversion compared to Alam's conversion was small.

The bias and LoA depended on the 3 T-R_2* to LIC method, as shown by the Bland Altman plots and Table 11. For d'Assignies' method the bias was -0.19 mg/g dw when comparing with Wood's 1.5 T-R_2* LIC and +0.35 mg/g dw when comparing with Hankins' 1.5 T-R_2* LIC. Apart from d'Assignies' 3 T-R_2* LIC, all other comparisons had a bias of less than ± 0.1 mg/g dw. Limits of agreement were generally within ± 0.2 mg/g dw, however the lower limit of agreement for d'Assignies' 3 T-R_2* LIC with Wood's 1.5 T-R_2* LIC was -0.55 mg/g dw and the upper limit of agreement for d'Assignies' 3 T-R_2* LIC with Hankins' 1.5 T-R_2* LIC was +0.64 mg/g dw. Using the method involving the 3 T to 1.5 T R_2* conversion equations found in this study (both

vendor specific and pooled) gave the smallest bias and tightest LoA. Using Alam's conversion did yield slightly higher biases in general than using the conversions generated from this study. However, the biases were still 0.08 mg/g dw or less. Also of note, the Bland Altman plots for d'Assignies (Figure 20) both show that there is a clear non-constant positive relationship between the bias and the magnitude of the paired values. This trend is not seen on the other Bland Altman plots.

Splitting the data depending on whether the 3 T scanner was Siemens (2.89 T), or Philips/GE generally made little difference on the regression lines, correlation, agreement, bias, and LoA, regardless of 3 T conversion method used. When considering the statistically significant ICCs, Philips/GE tended to have slightly higher ICCs, but by at most 0.2, when compared with the pooled data/Siemens ICCs. Similarly, using the specific R_2* conversions for Siemens/Philips/GE 3 T scanners found in this study made little difference to the correlation, agreement, bias, and LoA compared to when the pooled data R_2* conversion was used. For example, the improvement in bias for both Siemens and Philips was 0.02 mg/g dw or less, when compared to when a pooled conversion was used.

When comparing the impact of 1.5 T-R_2* LIC conversion method, Wood and Hankins both yielded similar correlation and agreement with 3 T-R_2* LIC (see Table 9 and Table 10). However, the scatter plots show that Hankins' 1.5 T-R_2* LICs are generally lower than Wood's 1.5 T-R_2* LICs (the data points are shifted to the left). When comparing with d'Assignies' 3 T-R_2* LIC the bias is quite different between the two and is less when Wood is used for 1.5 T-R_2* LIC compared to Hankins (-0.19 mg/g dw compared to 0.35 mg/g dw). For all the Bland Altman plots when Hankins was used for 1.5 T-R_2* LIC conversion there were broader LoA than for 1.5 T-R_2* LIC conversion there were broader

LoA than for Wood, however the difference was only small: generally, 0.02 mg/g dw or less.

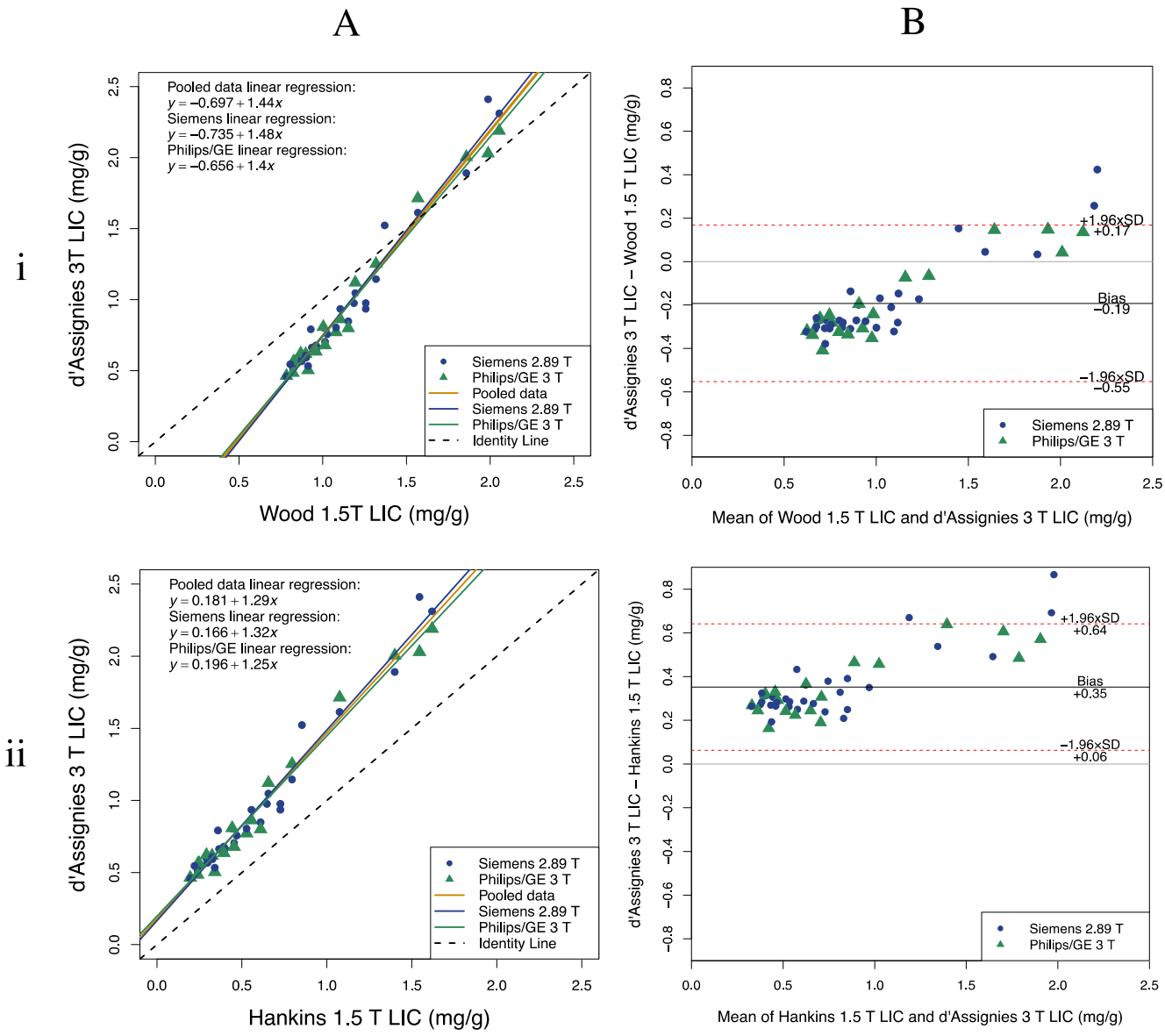


Figure 20: Scatter plots (A) and Bland Altman plots (B) comparing 3 T R_2* converted to LIC via d'Assignies and 1.5 T R_2* converted to LIC via (i) Wood and (ii) Hankins. Data point colour depends on the field strength (2.89 T or 3 T), and the three regression lines are for all the data, 2.89 T data and 3 T data.

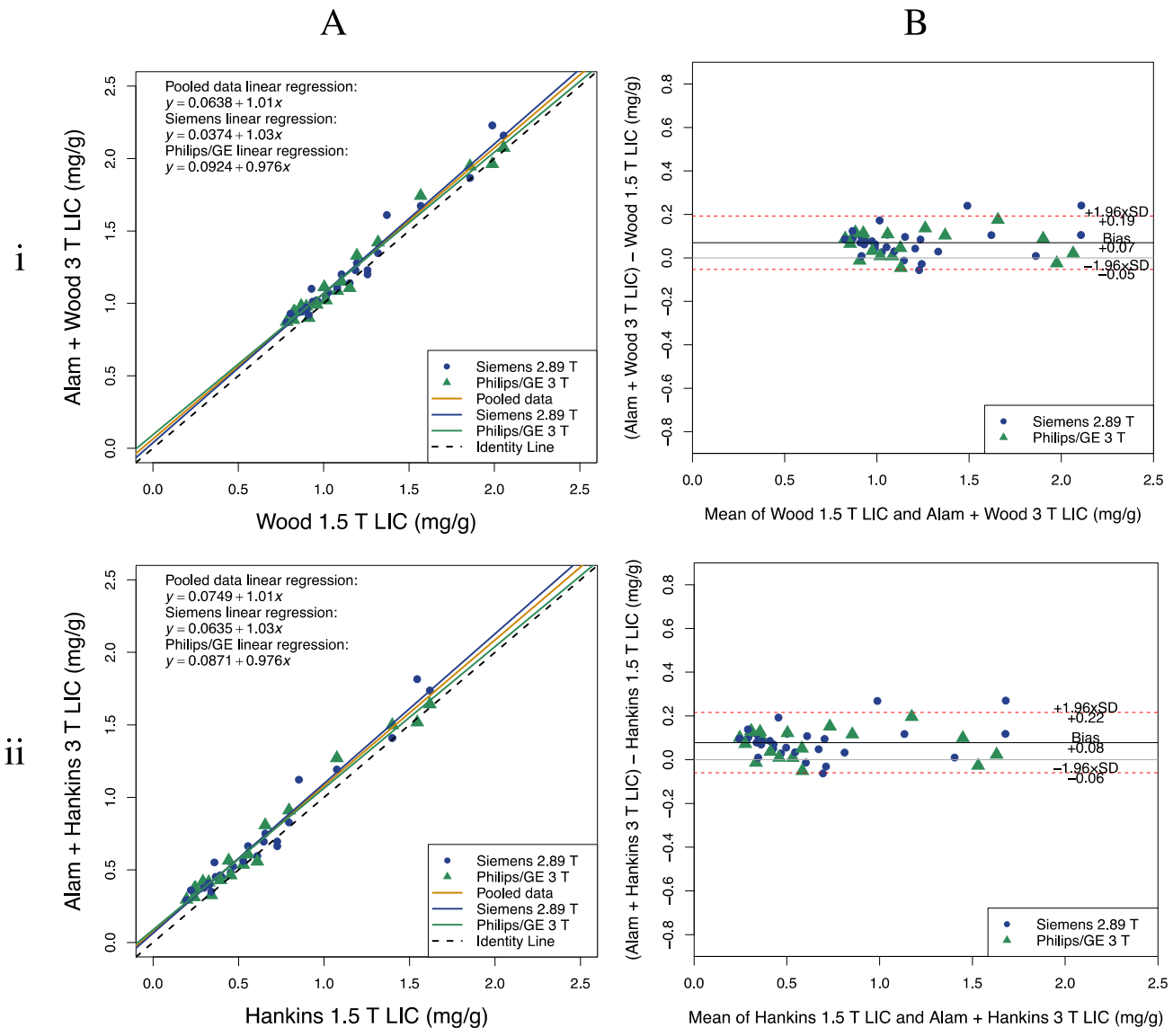


Figure 21 Scatter plots (A) and Bland Altman plots (B) comparing (i) 3 T R_2* converted to LIC via Alam and Wood and 1.5 T R_2* converted to LIC via Wood and (ii) 3 T R_2* converted to LIC via Alam and Hankins and 1.5 T R_2* converted to LIC via Hankins. Data point colour depends on the field strength (2.89 T or 3 T), and the three regression lines are for all the data, 2.89 T data and 3 T data.

	3 T Vendor	Number of samples	Wood 1.5 T LIC	p value	Hankins 1.5 T LIC	p value
d'Assignies 3 T LIC	All data	50	0.97	p <0.001	0.97	p <0.001
	Siemens	31	0.97	p <0.001	0.97	p <0.001
	Philips/GE	19	0.97	p <0.001	0.97	p <0.001
Alam R_2* Conversion with Wood/Hankins 3 T LIC *(Wood if Wood 1.5 T LIC or Hankins if Hankins 1.5 T LIC)*	All data	50	0.97	p <0.001	0.97	p <0.001
	Siemens	31	0.97	p <0.001	0.97	p <0.001
	Philips/GE	19	0.97	p <0.001	0.97	p <0.001
This study's pooled R_2* Conversion with Wood/Hankins 3 T LIC *(Wood if Wood 1.5 T LIC or Hankins if Hankins 1.5 T LIC)*	All data	50	0.97	p <0.001	0.97	p <0.001
	Siemens	31	0.97	p <0.001	0.97	p <0.001
	Philips/GE	19	0.97	p <0.001	0.97	p <0.001
This study's Siemens 2. 89 T R_2* Conversion with Wood/Hankins 3 T LIC *(Wood if Wood 1.5 T LIC or Hankins if Hankins 1.5 T LIC)*	Siemens	31	0.97	p <0.001	0.97	p <0.001
This study's Philips/GE 3 T R_2* Conversion with Wood/Hankins 3 T LIC *(Wood if Wood 1.5 T LIC or Hankins if Hankins 1.5 T LIC)*	Philips/GE	19	0.97	p <0.001	0.97	p <0.001

Table 9 Spearman's rho for 1.5 T-R_2* LIC and 3 T-R_2* LIC, for all data split depending on 3 T vendor

	3 T Vendor	Number of samples	Wood 1.5 T LIC	p value	Hankins 1.5 T LIC	p value
d'Assignies 3 T LIC	All data	50	0.84	0.004	0.74	0.060
	Siemens	31	0.82	0.006	0.72	0.063
	Philips/GE	19	0.87	0.004	0.78	0.055
Alam R_2* Conversion with Wood/Hankins 3 T LIC *(Wood if Wood 1.5 T LIC or Hankins if Hankins 1.5 T LIC)*	All data	50	0.97	0.001	0.97	0.001
	Siemens	31	0.96	0.001	0.96	0.001
	Philips/GE	19	0.98	p <0.001	0.98	p <0.001
This study's pooled R_2* Conversion with Wood/Hankins 3 T LIC *(Wood if Wood 1.5 T LIC or Hankins if Hankins 1.5 T LIC)*	All data	50	0.98	p <0.001	0.98	p <0.001
	Siemens	31	0.98	p <0.001	0.98	p <0.001
	Philips/GE	19	0.99	p <0.001	0.99	p <0.001
This study's Siemens 2. 89 T R_2* Conversion with Wood/Hankins 3 T LIC *(Wood if Wood 1.5 T LIC or Hankins if Hankins 1.5 T LIC)*	Siemens	31	0.98	p <0.001	0.98	p <0.001
This study's Philips/GE 3 T R_2* Conversion with Wood/Hankins 3 T LIC *(Wood if Wood 1.5 T LIC or Hankins if Hankins 1.5 T LIC)*	Philips/GE	19	0.99	p <0.001	0.99	p <0.001

Table 10 Intraclass correlation coefficients for 1.5 T-R_2* LIC and 3 T-R_2* LIC, for all data split depending on 3 T vendor

	3 T Vendor	Number of samples	Bias	Lower LoA	Upper LoA
d'Assignies 3 T LIC + Wood 1.5 T LIC	All data	50	-0.19	-0.55	0.17
	Siemens	31			
	Philips/GE	19			
d'Assignies 3 T LIC + Hankins 1.5 T LIC	All data	50	0.35	0.06	0.64
	Siemens	31			
	Philips/GE	19			
Alam R_2* Conversion with Wood 3 T LIC + Wood 1.5 T LIC	All data	50	0.07	-0.05	0.19
	Siemens	31			
	Philips/GE	19			
Alam R_2* Conversion with Hankins 3 T LIC + Hankins 1.5 T LIC	All data	50	0.08	-0.06	0.22
	Siemens	31			
	Philips/GE	19			
This study's pooled R_2* Conversion with Wood 3 T LIC + Wood 1.5 T LIC	All data	50	-0.02	-0.14	0.10
	Siemens	31			
	Philips/GE	19			
This study's pooled R_2* Conversion with Hankins 3 T LIC + Hankins 1.5 T LIC	All data	50	-0.02	-0.16	0.11
	Siemens	31			
	Philips/GE	19			
This study's Siemens 2.89 T R_2* Conversion with Wood 3 T LIC + Wood 1.5 T LIC	Siemens	31	0.01	-0.11	0.13
This study's Siemens 2.89 T R_2* Conversion with Hankins 3 T LIC + Hankins 1.5 T LIC	Siemens	31	0.01	-0.13	0.15
This study's Philips/GE 3 T R_2* Conversion with Wood 3 T LIC + Wood 1.5 T LIC	Philips/GE	19	-0.00	-0.12	0.12
This study's Philips/GE 3 T R_2* Conversion with Hankins 3 T LIC + Hankins 1.5 T LIC	Philips/GE	19	-0.00	-0.14	0.13

Table 11 Bland Altman parameters for 1.5 T-R_2* LIC and 3 T-R_2* LIC, for all data split depending on 3 T vendor

4.4. Discussion

It is important that iron concentration levels are estimated sufficiently accurately so that clinically significant changes in iron can be detected, whether that be for detecting iron sparing in the case of HCC, or for detecting and monitoring iron overload in patients with diseases such as haemochromatosis and thalassemia. LIC estimated from 1.5 T R_2* is widely used and is established as a method which has reliable precision and accuracy for clinical practice[40]. 3 T-R_2* LIC is not as widely used and there are fewer studies to verify its accuracy. In this study 3 T-R_2* LIC was compared with 1.5 T-R_2* LIC, to assess whether 3 T-R_2* LIC can be used by clinicians and researchers with the same confidence that 1.5 T-R_2* LIC is used. Additionally, comparing 3 T-R_2* LIC with 1.5 T-R_2* LIC enables the assessment of whether 1.5 T-R_2* LIC and 3 T-R_2* LIC can be used interchangeably, for example when a patient undergoing monitoring is scanned at both 1.5 T and 3 T or if a clinical trial is performed at both field strengths. 30 participants were scanned on multiple scanners to enable this comparison. 3 T-R_2* LIC was obtained using a number of different 3 T-R_2* to LIC conversion methods and was assessed against both Wood's[57] and Hankins'[58] 1.5 T-R_2* LIC.

One of the methods used for converting 3 T R_2* to LIC is the calibration curve from d'Assignies' study. This conversion yielded the poorest agreement with 1.5 T-R_2* LIC, which in some cases was not statistically significant. Additionally, the largest biases and widest LoA with 1.5 T-R_2* LIC were found, compared with the other 3 T-R_2* LIC conversion methods. Of particular concern is that the bias being greater than 0.23 mg/g dw when Hankins was used for 1.5 T-R_2* LIC, and the LoA being over ±0.23 mg/g dw when both Wood and Hankins were used for calculating 1.5 T-R_2* LIC. Additionally, 3 T-R_2* LIC for d'Assignies was often more than 0.23 mg/g dw from 1.5 T-R_2* LIC. From

the Bland Altman plots there also seemed to be an increasing positive difference between the 3 T-R_2* LIC and 1.5-R_2* LIC, which indicates that at higher iron levels the difference may be even greater. Therefore, it seems that it may not be appropriate to use this conversion for treatment monitoring and measurement of iron sparing. Given the trend towards increasing differences at higher LIC, it is probably not appropriate to use this method for diagnosis either.

It is not obvious why d'Assignies' 3 T-R_2* LIC conversion yields different results to Wood/Hankins' 1.5 T-R_2* LIC conversion. The calibration curves in both d'Assignies' and Wood/Hankins' studies were derived from R_2* data using ME sequences, like the one in this study, and from LIC obtained by biopsy. One reason could be the difference in patient populations, since the d'Assignies' curve was based on patients with LIC < 7.26 mg/g dw, whereas Wood and Hankins includes patients up to 32.9 mg/g dw [57] and 27.6 mg/g dw [58], respectively. However, both Wood and Hankins included patients with lower LIC in their studies. This was particularly so in the case of Hankins, where the lowest LIC was 0.6 mg/g dw [58]. Another reason could be differences in the biopsy methods used. Wood and Hankins do not give extensive details of the biopsy procedure itself, however it is possible that biopsies were performed differently compared to in d'Assignies' study. Additionally, the liver iron quantification for both Wood and Hankins studies was done at Mayo Medical Laboratory, Rochester, MN, whereas for the study by d'Assignies' *et al.* the quantification may have been done elsewhere, using a different technique. d'Assignies states that they use the Barry and Sherlock technique [78].

The other methods used to obtain 3 T-R_2* LIC all involved an initial step of converting from 3 T R_2* to 1.5 T R_2*, before converting to LIC using a 1.5 T-R_2* to LIC conversion method. Alam provided one method of 3 T R_2* to 1.5 T R_2* conversion in a previous study [60], but conversion equations were also generated using the data from this present

study. The conversion equations were similar but not identical. All the 3 T-R_2* LIC obtained using the two-step method, regardless of using Alam's or this study's R_2* conversion, showed excellent agreement and smaller biases with 1.5 T-R_2* LIC. Using Alam's conversion did yield slightly lower agreement and slightly higher biases in general than using the conversions generated from this study. This is to be expected as the conversion was not generated from the study data. However, given that the agreement was still excellent, and the biases were smaller than 0.1 mg/g dw, it seems that it would still be acceptable to use Alam's method for conversion from a clinical perspective of measuring iron sparing, along with monitoring of treatment for thalassemia when accuracy of 0.1 mg/g dw and 0.23 mg/g dw respectively is required.

Comparison was also made when data was split into Siemens 2.89 T and Philips/GE 3 T. However, there was very little difference in the linear regression equations on the graphs and the correlation and agreement, when compared to the pooled data results. This is perhaps to be expected when using d'Assignies for 3 T-R_2* LIC, as that study included both Siemens and Philips 3 T scanners[76]. Alam's study was on Siemens scanners, so it might be expected that the LIC calculated from the Siemens data would have the closest agreement with 1.5 T-R_2* LIC when using Alam's method, but this was not the case: Philip/GE had slightly higher agreement. Even so, the difference in agreement was minimal. Even when LIC was calculated using vendor-specific conversions from 3 T R_2* to 1.5 R_2*, the improvement in bias was minimal compared to when a pooled conversion was used. It should be noted that fewer data points were used to determine the vendor-specific conversions compared to the pooled conversions. Having more data points for each vendor may have created a vendor-specific conversion which yielded a more noticeable improvement compared to the pooled conversion. Given the excellent

agreement and small biases without the vendor-specific conversion, it seems that it is probably not essential to have different conversion methods depending on the 3 T vendor.

As well as different 3 T-R_2* LIC methods, two 1.5 T-R_2* LIC methods were used throughout. Interestingly, it is apparent from the scatter plots that Hankins' method generally yields lower 1.5 T-R_2* LIC. Although this was not the focus of this study, it is important that attention is drawn to this difference. Hankins' method also tended to yield slightly higher biases and LoA, which, although they are unlikely to have a clinical impact, may mean that it is advisable to use Wood's equation.

There are several avenues for further work. Firstly, a study with biopsy paired LIC would be preferable to using 1.5 T-R_2* LIC as the "ground truth". However, as 1.5 T-R_2* LIC is widely accepted, it would be hard to ethically justify performing biopsies when 1.5 T-R_2* LIC is available. Secondly, further studies are required on participants with higher LICs as this study only involved participants with LIC < 2 mg/g dw. However, it is applicable to HCC patients, who generally have low iron concentrations and even less iron in the HCC[17,19].

4.5. Summary

This study suggests that 3 T MRI can be used to give estimates of LIC with sufficient accuracy for clinical and research purposes, including in the context of detecting and measuring iron sparing, but care should be taken over which method is used to convert from R_2* to LIC. Given the findings from this work, the combination of Alam's and Hankins' equations were generally used in this project. Although the results with Hankins' equation were slightly worse than Wood's it was felt that Hankins' population was better matched to the patient population with HCC, hence the decision to use Hankins'.

5. Initial Evaluation of R_2* sequences: Phantom

5.1. Introduction

After establishing the requirements for an iron sparing sequence, I moved on to evaluating some existing sequences for 3D R_2* mapping, which could potentially be used in this application. Vendors such as Philips and Siemens have 3D ME GRE sequences which can be used to generate inline or offline T_2* and R_2* maps. I began by evaluating the results from the Philips mDixon Quant sequence with vendor parameters in a phantom, particularly focusing on assessment of concentrations where longer T_2*s/higher R_2*s were expected (i.e., equivalent to the lower iron expected in HCC patients). Phantoms are useful as we have a ground truth reference value to assess against, unlike *in vivo*. The scanning of the phantom was performed by my collaborators in Cincinnati; I undertook all the image analysis and statistical analysis.

After the initial evaluation of the phantom, I moved on to assessing the impact of using compressed sensing on R_2* estimates in the same phantom, again using the Philips mDixon Quant sequence. Compressed sensing can be used to accelerate the acquisition so could be useful in reducing the breath-hold length or for allowing for increased resolution. However, it is important to verify that it gives R_2* results in agreement with results without compressed sensing before it can be used in the clinic, hence the comparison with and without compressed sensing in the phantom. Again, the scanning of the phantom was performed by my collaborators in Cincinnati; I undertook all the image analysis and statistical analysis.

Although the aim of this work was initially to evaluate these sequences for the application of detecting iron sparing, it has wider applications and contributions to the clinical community. The vendor 3D ME GRE sequences are becoming increasingly

widespread and so it is important that they are evaluated against "gold standard" 2D ME GRE sequences before being used in the clinic. For example, it is important that the same diagnoses of iron overload category would be assigned if scanned with either sequence. Additionally, it is important that if 2D and 3D ME GRE sequences are used interchangeably for monitoring of iron during therapy in diseases such as thalassemia that the results are similar enough to be able to monitor the changes. Both of these applications are also discussed in this chapter.

5.2. Initial Evaluation using a Phantom

5.2.1. Introduction

Given the increasing use of 3D ME GRE for R_2* quantification in the liver, it is important to establish whether the results from this acquisition are comparable, in the clinical context, to "ground truth" values of R_2*. To this end, a validated R_2* phantom was imaged using 2D ME GRE (mFFE) and the (Philips Healthcare product) 3D ME GRE (mDixon Quant) sequences at 1.5 T. The corresponding R_2* and T_2* maps were then generated by inline and offline software algorithms. The results were then compared to the "ground truth" reference values, in the context of clinical thresholds. It is important that the differences between the results are such that the 3D sequences can be used with confidence alongside established sequences such as 2D ME GRE.

5.2.2. Method

A Calimetrix R_2* Phantom (Model 100) (Calimetrix, Madison, Wisconsin, United States) was used, with MR-verified R_2* values at 1.5 T. The vials within the phantom contain deionized water, agar, sodium chloride, sodium benzoate and manganese chloride tetrahydrate. Figure 22 shows a localizer image of the phantom.

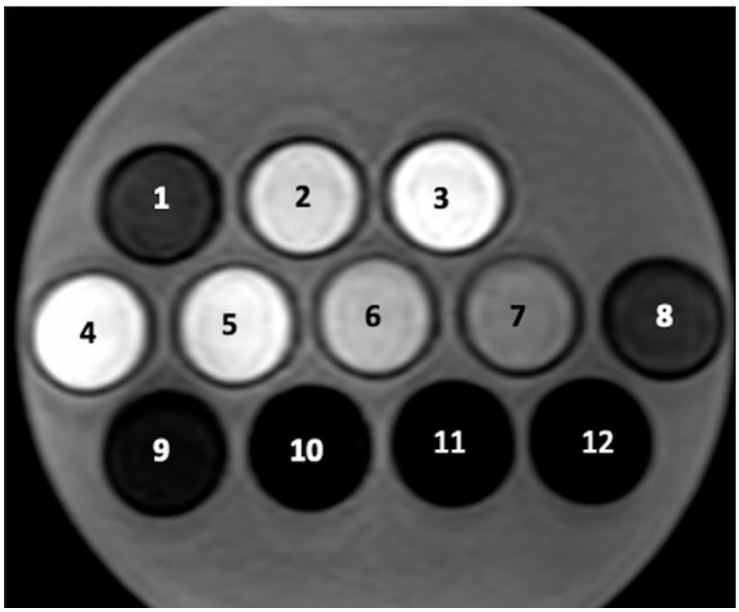


Figure 22 Localizer of phantom, with vials labelled 1-12

The 1.5 T R_2* reference values were obtained by Calimetrix by scanning the phantom using a 1.5 T Optima MR450 scanner (GE Healthcare, Waukesha, WI, United States). The IDEAL IQ sequence was used, with different parameters for vials 1-8 and 9-12. Table 12 gives the sequence parameters (in the final two columns). The reference values were assumed to be "ground truth" values, given that Calimetrix have optimised their sequences particularly for the R_2* values of the vials. However, it should be noted that the sequences used are 3D ME GRE sequences, similar to mDixon Quant (which was used in this study), but the processing was a simple exponential fitting, unlike the inline processing method for mDixon Quant, which uses a multi-peak fat model[79].

The corresponding LIC for each of the R_2* verified values were determined using the calibration curve reported by Hankins *et al.*[58] for 1.5 T and a liver iron overload classification assigned[54]. The verified R_2*values and equivalent LIC and iron overload assignment for each vial are reported in the first three columns of Table 13. Given that Vial 1 corresponded to a negative LIC value, it was excluded from all analyses.

For this study, the R_2* phantom was scanned using a 1.5 T Ingenia scanner (Philips Healthcare, Best, The Netherlands) with a 2D ME GRE (mFFE); and a 3D ME GRE (mDixon Quant). Both sequences were repeated 3 times. The mFFE is the standard 2D clinical sequence used at the medical centre where the phantom data was collected. The acquisition parameters for the scans are documented in Table 12. Inline R_2* and T_2* maps were generated on the MR scanner. Offline R_2* and T_2* maps were generated using the MAGO technique[65], as described in Section 3.7.

Parameter	Philips 2D ME GRE (mFFE)	Philips 3D ME GRE (mDixon Quant)	GE 3D ME GRE Calimetrix IDEAL IQ vials 1 - 8	GE 3D ME GRE Calimetrix IDEAL IQ vials 9 - 12
Field of View (mm x mm)	350 x 350	400 x 350	240 x 240	240 x 240
Acquisition Matrix size	140 x 140	132 x 118	144 x 144	144 x 144
Acquired Resolution (mm)	2.50 x 2.50	3.03 x 2.97	1.67 x 1.67	1.67 x 1.67
Reconstructed voxel size (mm)	0.99 x 0.99	2.08 x 2.08	0.94 x 0.94	0.94 x 0.94
Acquired slice thickness (mm)	5.00	6.00	3.00	3.00
Reconstructed slice thickness (mm)	5.00	3.00	3.00	3.00
TR (ms)	15.02	5.32	24.88	11.55
Number of echoes	16	6	6	4
Initial TE (ms)	1.12	0.92	1.14	0.80
Final TE (ms)	13.80	4.22	11.16	6.29
Echo spacing (ms)	0.85	0.66	2.00	0.78
Flip Angle (degrees)	25	5	3	8

Table 12 Scanner Parameters, for the Philips acquisitions used in this study: 2D and 3D multi-echo gradient recalled echo (ME-GRE) and for the reference values obtained by Calimetrix on a GE scanner, with two different sets of parameters for vials 1-8 and 9-12

Although multiple slices were acquired for both sequences, to be consistent with clinical practice, only one slice was used for subsequent analysis. The slice selected was the one placed at the centre of the phantom, and this was consistent across the acquisitions, although the level differed by up to 1 mm due to the slice locations not being identical between the 2D and 3D sequences. 15 mm circular ROIs were placed centrally on each vial and the median value of the pixels was recorded. OsiriX Lite (Pixmeo SARL, CH-1233 Bernex, Switzerland) was used to place analogous ROIs on the inline maps. In-

house software was used to place ROIs on the offline maps. The mean and standard deviation of the ROI R_2* medians from the three repetitions of each sequence were calculated and recorded. A scatter plot of the results visually compared the results from each ROI to the reference values. As in the previous chapter, Spearman's rho (correlation), ICC (using the first of the three repetitions of R_2* results for each, as opposed to the mean of the three repetitions of each sequence) were used for statistical assessment, alongside Bland Altman plots. p values of < 0.05 were classed as statistically significant. Repeatability was assessed with scatter plots, Spearman's rho, ICC, and Bland-Altman plots, using the first two of the three repetitions for each sequence and processing method.

Similar to the study in Chapter 4, it is important to consider the thresholds of permissible variation between the ground truth and different sequences in light of a variety of clinical applications. For diagnosis of iron overload, there are classification categories are available, as given in Chapter 4: 3.2 – 7.0 mg/g dw is considered mild overload, 7.0 – 15.0 mg/g dw is moderate overload, while >15.0 mg/g dw is severe[56]. Although the phantom does not contain iron, converting the R_2* to the equivalent LIC makes the evaluation of the phantom clinically relevant. Therefore, in this study, the R_2* values were converted to LIC and categorised according to the above classification, to determine if the values obtained by the sequences under analysis were the same as those for the reference values.

For the purpose of monitoring patients over time, for example patients being treated with iron chelation in diseases such as thalassemia, minimising between-sequence variability is important if different sequences or processing methods are to be used interchangeably in monitoring. The acceptable variation of R_2* is determined by the amount of variation that needs to be detected. As described in Chapter 4, using the results

from the study measuring the impact of the iron chelator Deferasirox, it is desirable to be able to detect a change of 0.45 mg/g dw over a 12 week period[77]. To accurately detect such a change, differences in R_2* between acquisitions need to correspond to less than 0.23 mg/g dw. Therefore using 0.2 mg/g dw as the threshold this means R_2* variability needs to be less than 7.14 s^{-1}.

To detect iron sparing, as specified in Section 3.8.4, we have the requirement to measure iron concentration to within 0.1 mg/g dw. Therefore, at most R_2* differences corresponding up to 0.1 mg/g dw are acceptable, which corresponds to 3.57 s^{-1}.

5.2.3. Results

The means and standard deviations of the three repeats are reported in Table 13. The scatter plot comparing the reference values and the R_2* estimates produced by the 2D mFFE and 3D mDixon Quant sequences are shown in Figure 23. Table 14 and Table 15 give the Spearman's rho and ICC for assessing the correlation and agreement between the different measurements and with the reference values. Scatter plots and Bland-Altman plots comparing the first acquisition and the first repetition of each sequence are shown in Figure 24, with Spearman's rho and ICC for the first two repeats are shown Table 16.

				Offline processing								Inline processing							
Vial	**Calimetrix MR-verified R_2* value from user manual (s^{-1})**	**Calimetrix MR-verified R_2* value from user manual converted to LIC (mg/g dw) (Iron overload category)**		**mFFE R_2* Mean from 3 repeats (s^{-1}) (Standard Deviation)**		**mFFE R_2* Mean converted LIC (mg/g dw) (iron overload category)**		**mDixon R_2* Mean from 3 repeats (s^{-1}) (Standard Deviation)**		**mDixon R_2* Mean converted to LIC (mg/g dw) (iron overload category)**		**mFFE R_2* Mean from 3 repeats (s^{-1}) (Standard Deviation)**		**mFFE R_2* Mean from 3 repeats converted to LIC (mg/g dw) (iron overload category)**		**mDixon R_2* Mean from 3 repeats (s^{-1}) (Standard Deviation)**		**mDixon R_2* Mean from 3 repeats converted to LIC (mg/g dw) (iron overload category)**	
1	12	-0.12	(N/A)																
2	24	0.22	(Normal)	26.11	(0.39)	0.28	(Normal)	12.94	(0.37)	-0.09	(N/A)	25.09	(0.59)	0.25	(Normal)	35.03	(3.34)	0.53	(Normal)
3	55	1.09	(Normal)	67.56	(0.51)	1.44	(Normal)	71.33	(4.21)	1.54	(Normal)	67.70	(0.39)	1.44	(Normal)	80.12	(2.11)	1.79	(Normal)
4	81	1.81	(Insignificant)	91.29	(0.17)	2.10	(Insignificant)	118.57	(1.07)	2.86	(Insignificant)	90.05	(0.23)	2.07	(Insignificant)	124.08	(0.00)	3.02	(Insignificant)
5	104	2.46	(Insignificant)	114.73	(0.20)	2.76	(Insignificant)	46.05	(0.40)	0.83	(Normal)	113.39	(0.36)	2.72	(Insignificant)	119.86	(0.41)	2.90	(Insignificant)
6	156	3.91	(Mild)	162.81	(0.72)	4.10	(Mild)	121.98	(17.93)	2.96	(Insignificant)	162.45	(0.76)	4.09	(Mild)	152.82	(1.98)	3.82	(Mild)
7	212	5.48	(Mild)	211.58	(0.20)	5.47	(Mild)	184.45	(1.44)	4.71	(Mild)	212.36	(0.00)	5.49	(Mild)	203.08	(2.32)	5.23	(Mild)
8	319	8.48	(Moderate)	311.21	(10.52)	8.26	(Moderate)	321.46	(2.58)	8.55	(Moderate)	324.57	(0.00)	8.63	(Moderate)	228.20	(9.75)	5.93	(Mild)
9	422	11.36	(Moderate)	417.17	(2.54)	11.23	(Moderate)	368.95	(1.66)	9.88	(Moderate)	418.41	(0.00)	11.26	(Moderate)	415.01	(2.49)	11.17	(Moderate)
10	640	17.47	(Severe)	624.76	(6.74)	17.04	(Severe)	581.06	(2.92)	15.81	(Severe)	644.33	(0.00)	17.59	(Severe)	614.28	(5.25)	16.74	(Severe)
11	853	23.43	(Severe)	775.21	(9.37)	21.25	(Severe)	814.11	(6.87)	22.34	(Severe)	864.30	(0.00)	23.75	(Severe)	782.53	(8.43)	21.46	(Severe)
12	1070	29.51	(Severe)	905.05	(24.99)	24.89	(Severe)	943.14	(17.54)	25.95	(Severe)	1100.67	(59.91)	30.36	(Severe)	890.47	(0.00)	24.48	(Severe)

Table 13 Mean R_2* values of three repeats of sequence for each vial (excluding Vial 1), converted to, converted to LIC using Hankins' equation[58] and categorised in terms of iron overload[56]

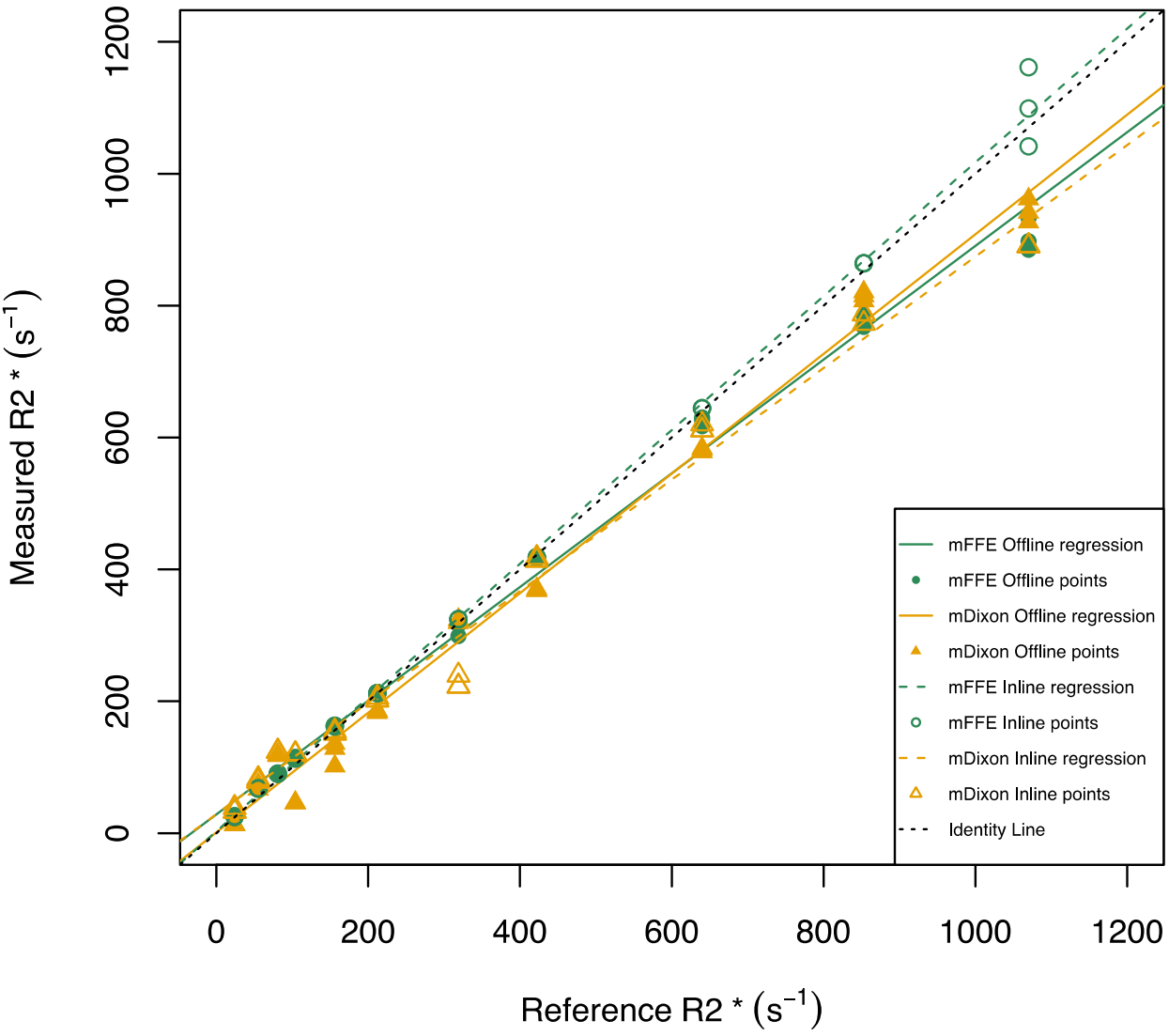


Figure 23 Scatter plot comparing the phantom reference values with the measured values, with lines of best fit (linear regression)

	mFFE Offline	mDixon Offline	mFFE Inline	mDixon Inline	Verified values
mFFE Offline					
mDixon Offline	0.95 ($p < 0.001$)				
mFFE Inline	1.00 ($p < 0.001$)	0.95 ($p < 0.001$)			
mDixon Inline	0.99 ($p < 0.001$)	0.98 ($p < 0.001$)	0.99 ($p < 0.001$)		
Verified values	1.00 ($p < 0.001$)	0.95 ($p < 0.001$)	1.00 ($p < 0.001$)	0.99 ($p < 0.001$)	

Table 14 Spearman's rho values for comparing each of the measured values from the first acquisition with each other, and with the reference values

	mFFE Offline	mDixon Offline	mFFE Inline	mDixon Inline	Verified values
mFFE Offline					
mDixon Offline	0.99 ($p < 0.001$)				
mFFE Inline	0.98 ($p < 0.001$)	0.97 ($p < 0.001$)			
mDixon Inline	>0.99 ($p < 0.001$)	0.99 ($p < 0.001$)	0.96 ($p < 0.001$)		
Verified values	0.99 ($p < 0.001$)	0.98 ($p < 0.001$)	>0.99 ($p < 0.001$)	0.98 ($p < 0.001$)	

Table 15 ICC values for comparing each of the measured values from the first acquisition with each other, and with the reference values

	Spearman's rho	ICC
mFFE Offline	1.00 ($p < 0.001$)	>0.99 ($p < 0.001$)
mDixon Offline	0.99 ($p < 0.001$)	>0.99 ($p < 0.001$)
mFFE Inline	1.00 ($p < 0.001$)	>0.99 ($p < 0.001$)
mDixon Inline	1.00 ($p < 0.001$)	>0.99 ($p < 0.001$)

Table 16 Spearman's rho and Intra-class correlation coefficients for two repetitions

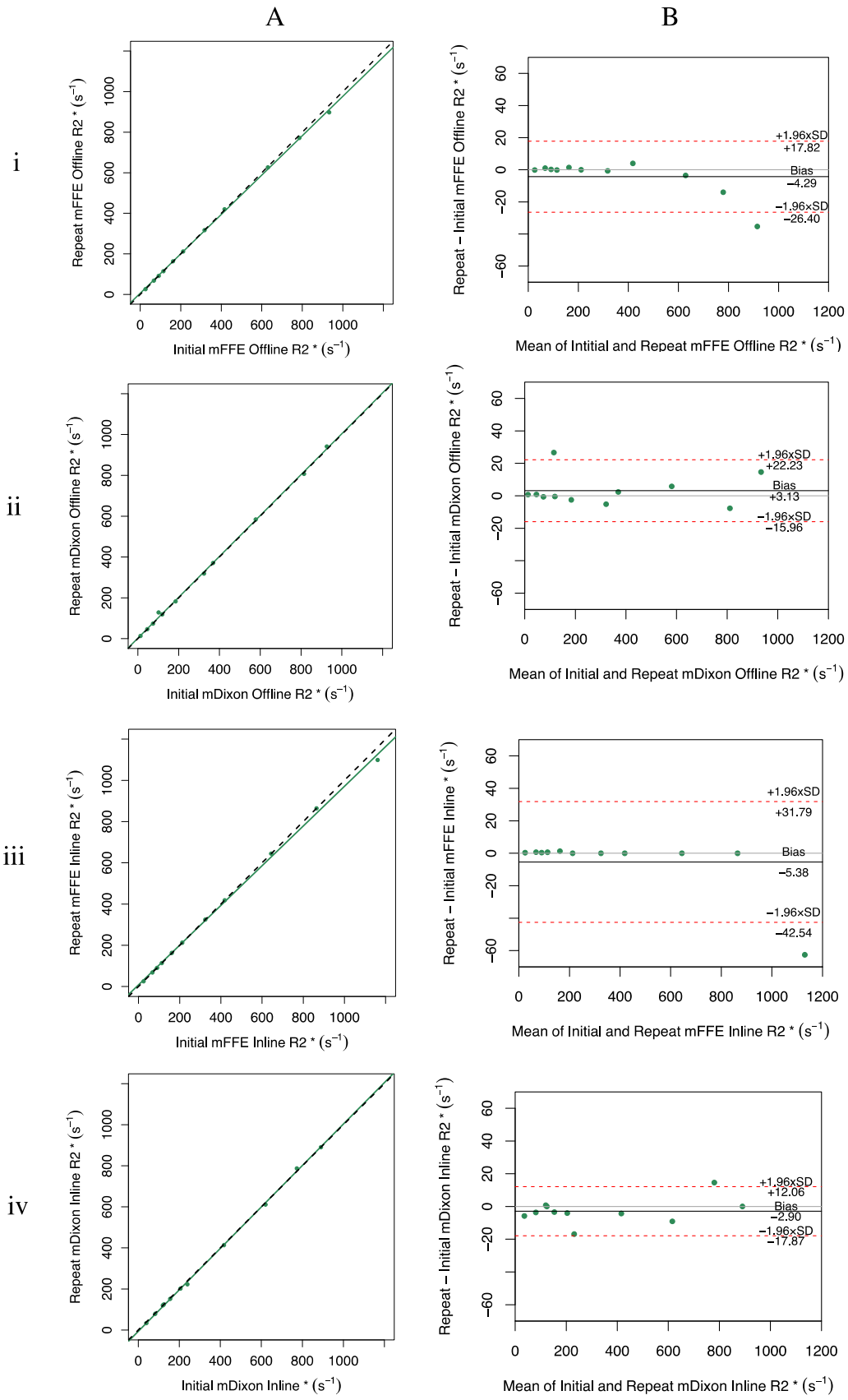


Figure 24 Scatter plots (A) and Bland Altman plots (B) comparing the first acquisition and first repeat for (i) mFFE offline, (ii) mDixon offline, (iii) mFFE inline and (iv) mDixon inline. For the scatter plots the dashed black line is the identity line and the solid green line is the line of best fit.

Table 14 and Table 15 show that there was a strong significant correlation and agreement between all sequences and the reference values. However, Table 13 shows that the R_2* values did vary between acquisitions and with the reference values, in some cases by up to 50 s^{-1}. However, the amount of variation between R_2* depended on sequence type, processing type and vial. As shown in Figure 23, the linear regression line for mFFE inline processing is the closest to the identity line (when comparing the measured values with the reference values), indicating that the results from mFFE inline processing were most similar to the reference values. However, even for mFFE inline processing, the difference between the mean and the reference value was over 30 s^{-1} for vial 12, although for the other vials the difference was much lower. In Figure 23 the other three regression lines are all close together, indicating similarity between the results from these processing methods.

Of most importance is the clinical impact of any differences between the sequences. As can be seen from Table 13 for mFFE inline and mFFE offline processing, the iron overload categorisation is the same as for the reference values for all vials. For mDixon inline the results are also similar: there is just one difference in classification (mild instead of moderate). However, for mDixon offline processing there are more differences, but for the final 6 vials, which are arguably the most important clinically, representing more advanced iron overload, the classification is consistent with that of the reference values.

However, when considering the use of these sequences for monitoring treatment such as chelation, and for detection of changes associated with iron sparing, the differences in iron concentration are greater than the thresholds specified of 0.2 mg/g dw and 0.1 mg/g dw across the majority of sequences and acquisitions. The differences are especially large between the acquisitions and with the reference values for the highest R_2* vials, 11 and 12, for example for mDixon inline processing the equivalent iron concentration was 5

mg/g dw less than the reference values. However, although the differences were largest for higher R_2*, even for lower R_2* the differences were still often outside of the 0.1 mg/g dw threshold, for example for vial 3 the differences to the reference value iron concentrations ranged from 0.35 mg/g dw to 0.70 mg/g dw.

mDixon Quant offline and mDixon inline had some of the largest differences in iron concentrations to the reference values, including six vials with a difference in R_2* equivalent to 1 mg/g dw or more for the offline processing and 4 for the inline processing. mFFE inline processing had the smallest differences in iron concentration, with no differences of more than 1 mg/g dw.

Generally, R_2* (and equivalent LIC) was higher for the measured values compared to the reference for lower R_2* values (R2* below 104 s^{-1}), whereas for higher R_2* values (R2* above 422 s^{-1}) (corresponding to higher LIC) the measured values were generally less than the reference values. The differences are not only between the sequences and the reference values, but also between each other. For vials 2-9 for mFFE online and offline processing the results have less than 0.1 mg/g dw difference, however for mDixon Quant the differences with mFFE and inline and offline processing, are often greater than the 0.1 mg/g dw and 0.2 mg/g dw thresholds.

In terms of repeatability, Table 16 shows that the correlation and agreement between the first acquisition and the repeat was strong and significant for all sequences and processing methods. However, the Bland Altman plots in Figure 24 show that the bias was in some cases more than the 0.1 mg/g dw threshold and the LoA were wider than the 0.1 mg/g dw and 0.2 mg/g dw R_2* thresholds. Additionally, the standard deviations, as shown in Table 13, were also sometimes greater than the R_2* equivalent to 0.1 mg/g dw and 0.2 mg/g dw iron concentration difference, although this did vary between vials. For example, vial 12 had a large standard deviation (>15 s-1) for most of the sequences and

processing methods, and vial 6 had a large standard deviation (>15 s-1) for mDixon offline processing. Given that the large standard deviation for vial 6 was confined to mDixon offline processing this was investigated further and was found to be due to the incorrect solution being selected for two of the acquisitions: i.e., the solution for the pixels in that area being selected as having 100% fat, when it actually had 0% fat.

For 3D mDixon Quant inline processing there was an artifact that was visible on the T_2* maps impacting some of the vials. The artifact was manifested as a marked difference in T_2* in a region of the phantom. This artifact was also present on inline generated PDFF maps. The impact on Vial 1 was significant, but this had no impact on the overall results, as vial 1 was excluded from all plots and statistical analyses since the LIC value was not realistic (it was negative). Vial 4 was also partially impacted by the artifact but, given that the median value was taken, the ROI result was not adversely affected.

5.2.4. Discussion

2D ME GRE R_2* mapping sequences are in widespread clinical use as a surrogate for liver iron concentration estimation, but 3D sequences are becoming increasingly popular and could be useful for identifying heterogeneity in iron throughout the liver, for example in iron sparing in HCC. This R_2* phantom study investigated whether both 2D and 3D acquisitions meet the clinical requirements for a sequence used to detect iron, including for iron overload diagnosis, treatment monitoring and detection of iron sparing. The phantom used did not contain iron but contained manganese chloride to create differences in R_2*.

Although the results overall showed strong correlation and agreement with each other and the reference values, the actual values could be quite different. This was the case throughout the different vials, although the differences to the reference values and to each other were largest for the higher R_2* vials. However, in the clinical context of detecting

iron overload these differences are not overly concerning, given that the categories for the highest iron overload were consistent with those for the reference values, across all sequences and processing methods.

However, the differences between sequences and with the reference values is more concerning in the context of monitoring treatment or the amount of iron sparing. Given the differences with the reference values and between sequences were often larger than R_2* equivalent 0.2 mg/g dw differences this means that if a patient was being monitored with different sequences "real" changes in iron might be masked/enhanced by differences in R_2* due to differences in the sequences. The particularly large differences at higher R_2* is concerning if patients with iron overload are being monitored. However, even at low iron concentration, the differences are such that monitoring for iron sparing would be difficult between sequences. The differences between the sequences could be due to differences in sequence parameters, such as the number of echoes and the echo spacing. It may be desirable to have different parameters for different applications: for example, shorter first TE and echo spacing for detecting higher R_2* and a longer first TE and wider echo spacing for detecting changes in iron for iron sparing. The reference values were estimated using different sequences and so better accommodated the different R_2* values of the vials.

The different sequences had different amounts of agreement with the ground truth reference values. The smallest R_2* differences (i.e., equivalent to smallest LIC differences) were seen on the mFFE inline processing. This is not surprising for vials 2-8, given the similarity between the sequence parameters with the reference sequence. It is more surprising for vials 9-12, where the first TE was later than the reference sequence, which means it would be more challenging to accurately fit the rapid T_2* decay curve. However, the high flip angle and number of echoes may have helped to maintain a good

fit, and hence give comparable results to the reference. The mDixon Quant had larger iron concentration differences to the reference values for both inline and offline processing, and also differed from mFFE. This may have been due to the larger voxel size, making partial voluming and susceptibility more likely to impact the measurements.

Both the offline and inline processing of mDixon Quant resulted in some errors in the T_2* and R_2* maps, For the offline processing this was because the incorrect solution for some pixels was selected (100% PDFF instead of 0% PDFF), which is unlikely to happen *in vivo*, where there will always be some fat. For the inline processing the artifact was likely due to the inline algorithm containing assumptions about the body which are violated in the phantom. This highlights a limitation of processing techniques inline and offline when using a phantom.

The differences in the repeats of the sequence across all acquisitions and processing is concerning, especially for the highest R_2* values, as this indicates that, even if the same sequence is used for monitoring, non-physiological differences in R_2* could mask/enhance the estimated iron difference. However, this may not necessarily be the case *in vivo*.

Previous studies have compared 2D ME GRE sequences with 3D ME GRE sequences in populations of patients, with a variety of background aetiologies[80–82], at 1.5T. They have shown good agreements between the two sequences. For example, Serai *et al.*[82] compared a 2D ME GRE sequence with 3D mDixon Quant on Philips 1.5 T scanners in a paediatric population. The sequences and acquisition parameters for both the 2D ME GRE and 3D mDixon Quant sequences were similar to those used in this study, with the exception of the number of echoes for the 2D ME GRE sequence (8 for their sequence, compared to 16, with their final TE being 6.6 ms versus 13.803 ms). Although they showed excellent absolute agreement between the two sequences, six patients did fall

outside the 95% LoA, all of whom had T_2* greater than 20 ms. Their 2D GRE ME sequence collected more echoes, and a later final TE compared to the 3D mDixon Quant and this may be the reason why some of the patients with longer T_2* estimates were outside the LoA. Although there was excellent agreement in their Bland-Altman plots, they showed a mean bias of 0.10 ms with 1.96 SD limits of −4.40 to +4.60 ms. Consequently, in these cases, the two techniques may have assigned a given patient to different liver iron categories. Thus, overall, their findings are for the most part consistent with ours: specifically, although there is a correlation between the sequences, there are also substantial differences between them.

Similarly, Henninger *et al*. compared 3D ME GRE sequences to a 2D ME GRE sequence on Siemens 1.5T scanners in two separate studies[80,81]. Notably, the 3D ME Dixon sequences used were different between the two studies, and in the second study two different 3D sequences were used, qDixon and a qDixon works in progress (WIP) version. In both studies the authors found good agreement between the 2D and 3D methods.

In the first study, Henninger *et al.*[80] found a systematic under-estimation of R_2* by the 3D compared to the 2D method. However, they did not find the same to be true in the second study[81] in which the qDixon WIP sequence, for which the parameters and post processing were different, was used. Additionally, in the first study, 14 out of 120 cases were discarded due to fat-water swaps or because they could not report on R_2* values above $400s^{-1}$ (i.e., very short T_2* values). However, these issues were obviated in the second study through changes to the fitting and post processing techniques. Cases like these highlight the need for and/or potential advantages/benefits of the availability/application of offline fitting approaches. Using the updated sequence and post processing method in the second paper, Henninger *et al.*[81]. reported a mean bias of -3.82

s^{-1}, but with points with higher R_2* showing larger differences between the techniques, which is consistent with our study where the biggest differences in R_2* were those vials with the higher R_2*. This is probably because the shorter T_2* is challenging to estimate given the rapid signal decay for these species compared to the relatively long first echo time and inter echo spacings employed by standard T_2* 2D and 3D.

In Henninger's second study using the updated 3D ME GRE sequence they compared the results of different calibration methods for categorisation of patients using qDixon. They found the overall agreement for classification was >90% in most cases, with the different methods at most misclassifying by one category[81]. This is similar to our findings on iron classification.

Our study had a number of limitations. First, the phantom did not contain iron and so did not truly represent iron in the liver. However, creating a phantom which truly represents iron in the liver is challenging. For example, super-paramagnetic nanoparticles of iron oxide can be used but the structure of the iron does not match that of iron in the liver[83]. Using iron oxide microspheres have shown some promising results[84] but there is no commercially available phantom (to our knowledge) which uses iron oxide microspheres. It is also true that fat can influence R_2*[85] and should ideally be included as well, but again, although there are a few studies where fat has been included within R_2* phantoms[84,86], these are not (to our knowledge) commercially available.

Another limitation is that the study primarily compared the results with reference results, which could, in themselves, not be accurate. However, given the sequences were designed and run by a company who specialises in phantoms, it can be assumed that they endeavoured to make sure the results were as accurate as possible, by optimising the sequences and choosing fitting methods which best work for the R_2* values and phantom composition.

Another limitation was that the data was only analysed from a 1.5 T scanner. In future work the experiment could be repeated with data from a 3 T scanner.

Finally, the acquisition parameters used to collect the 2D mFFE and 3D mDixon Quant imaging data were not identical, and so 'like is not compared with like', but instead used manufacturer-suggested clinically relevant parameters for each sequence under study.

5.3. Evaluation of Compressed Sensing using a Phantom

This subsection covers a project comparing results with and without compressed sensing, using the same phantom as used in Section 5.2.

5.3.1. Introduction

3D R_2* mapping sequences are usually acquired in a breath-hold, but these can be lengthy, particularly if high resolution or long echo times are used. Long breath-holds can be challenging for patients and so acceleration techniques are used to reduce the breath-hold length. For example, parallel imaging techniques such as Sensitivity Encoding (SENSE) can be used to accelerate acquisition (see Section 3.5.2), but the required breath-hold can still be quite long. Compressed sensing is a more recent development in MRI (see Section 3.5.3) and can be used to further accelerate the 3D ME GRE acquisition. However, before compressed sensing can be used in the clinic on 3D R_2* mapping we need to demonstrate sufficient similarity in the results obtained with and without compressed sensing. Not only is this important for initial diagnosis, but also for monitoring, particularly if a patient is scanned with and without compressed sensing at different time points. Use of acceleration is of particular importance if using the 3D ME GRE for detection of iron sparing, given the high resolution and longer TEs required lead to a long breath-hold. Therefore, this work has relevance to HCC detection. In this study

I compared the R_2* estimates obtained in a validated phantom using a 3D ME GRE acquisition with and without compressed sensing.

5.3.2. Method

The Calimetrix R_2* phantom, also used in Section 5.2, containing 12 vials of varying R_2* was scanned on a 1.5 T Philips Ingenia MRI scanner using an mDixon Quant sequence. The sequence was first performed with SENSE (the standard parallel imaging technique on Philips scanners) factor 2, and then with Compressed SENSE (the compressed sensing product on Philips scanners) factor 2, 4 and 6. The sequence parameters are given in Table 17.

	mDixon Quant SENSE 2	**mDixon Quant Compressed SENSE 2**	**mDixon Quant Compressed SENSE 4**	**mDixon Quant Compressed SENSE 6**
Field of View (mm x mm)	199 x 174	199 x 174	199 x 174	199 x 174
Acquisition Matrix size	132 x 112	132 x 111	132 x 111	132 x 111
Acquired resolution (mm)	1.51 x 1.56	1.51 x 1.57	1.51 x 1.57	1.51 x 1.57
Reconstructed resolution (mm)	0.78 x 0.78	0.78 x 0.78	0.78 x 0.78	0.78 x 0.78
Acquired slice thickness (mm)	3.00	3.00	3.00	3.00
Reconstructed slice thickness (mm)	1.50	1.50	1.50	1.50
TR (ms)	7.43	7.43	7.43	7.43
Number of echoes	6	6	6	6
Initial TE (ms)	1.24	1.24	1.24	1.24
Final TE (ms)	6.14	6.14	6.14	6.14
Echo spacing (ms)	0.98	0.98	0.98	0.98
Flip Angle (degrees)	5	5	5	5
SENSE factor	2	N/A	N/A	N/A
Compressed SENSE factor	N/A	2	4	6
Acquisition Duration	12.58	14.11	7.29	4.96

Table 17 Sequence parameters

T_2* and R_2* maps were generated offline using the MAGO method[65]. 15 mm diameter circular ROIs were placed on each vial (excluding vial 1, which was outside of the physiological range of T_2* found in the liver). T_2* was also used for Compressed SENSE 6 to enable easier comparison with a previous *in vivo* study comparing SENSE and Compressed SENSE 6[87]. Median R_2* and T_2* was calculated for each ROI. Each acquisition was repeated to enable assessment of repeatability. As in previous sections, scatter plots and Bland Altman plots were made. Statistical tests included Spearman's rho (correlation), ICC (agreement) and Bland Altman statistics (bias and LoA). p values of < 0.05 were classed as statistically significant.

Results were assessed in a variety of clinical contexts, considering how much difference is acceptable for diagnosis of iron overload, monitoring treatment of iron overload and for detecting iron sparing. These were the same assessments as for Section 5.2, although the comparison was between 3D acquisitions with and without compressed sensing in this scenario. For the detection of iron overload, R_2* was converted to LIC using Hankins' equation[58] and the LIC classified as normal, insignificant, mild, moderate or severe, based on clinical categories[56]. For treatment monitoring differences equivalent to 0.2 mg/g dw are acceptable, this means R_2* variability needs to be less than 7.14 s^{-1} (using Hankins' conversion[58]). To detect iron sparing, as specified in Section 3.8.4, we have the requirement to measure iron concentration to within 0.1 mg/g dw. Therefore, at most R_2* differences corresponding up to 0.1 mg/g dw are acceptable, which corresponds to 3.57 $s^{-1.}$ (using Hankins' conversion[58]). If the differences between SENSE and Compressed SENSE are less than these thresholds, then this indicates that SENSE and Compressed SENSE can be used interchangeably.

5.3.3. Results

The results for R_2* are shown in Table 18 and Table 19, along with the equivalent iron concentration and its iron overload classification. The iron overload classification for each vial was the same between SENSE and Compressed SENSE (for all three acceleration factors).

Scatter plots and Bland-Altman plots comparing the initial SENSE and Compressed SENSE R_2* results are shown in Figure 25. Table 20 gives Spearman's rho and ICCs for comparing Compressed SENSE results with SENSE. There was strong and significant correlation and agreement between SENSE and Compressed SENSE for all Compressed SENSE results. However, for R_2* the Bland Altman plots (Figure 25) shows a large bias for both Compressed SENSE 2 (+16.75 s^{-1}) and Compressed SENSE 6 (-21.49 s^{-1}). For Compressed SENSE 4 the bias is less (-2.97 s^{-1}), and around midway between Compressed SENSE 2 and Compressed SENSE 6 bias. The 95% LoA were wide for all three Compressed SENSE factors. This indicates that for some of the vials the differences in R_2* are greater than the clinical thresholds for treatment monitoring and detecting iron sparing. Looking more closely as the Bland Altman plots, it is the 4 vials with the highest R_2* that have the largest differences in R_2* between SENSE and Compressed SENSE. For Compressed SENSE 2, the Compressed SENSE R_2* results for these vials were higher than SENSE, for Compressed SENSE 4 they were a mixture of higher and lower and for Compressed SENSE 6 R_2* estimates were all lower than for SENSE. In the 7 vials with smaller differences for all three Compressed SENSE factors there were instances where R_2* was greater than the 3.57 s^{-1} iron sparing threshold, but the differences are all less than the 7.14 s^{-1} threshold. Bland Altman statistics were calculated with only the seven vials with lowest R2*. Biases were less than when all vials were included (biases with seven vials: Compressed SENSE 2: -0.63 s^{-1}, Compressed SENSE

4: +2.81 s^{-1}, and Compressed SENSE 6: -0.80 s^{-1}) and LoA smaller (LoA with seven vials: Compressed SENSE 2: -5.08 s^{-1} and +3.83 s^{-1}, Compressed SENSE 4: -2.28 s^{-1} and +7.89 s^{-1}, and Compressed SENSE 6: -8.10 s^{-1} and +6.50 s^{-1}).

Figure 26 shows the scatter plot and Bland Altman plot comparing SENSE and Compressed SENSE 6 T_2* measurements. Spearman's rho was 1.00 ($p < 0.001$) and ICC was >0.99 ($p < 0.001$). The bias was small (-0.17 ms) with tight 95% LoA (+1.17 ms and -1.50 ms). The largest differences between SENSE and Compressed SENSE were found for vials with longer T_2*.

When comparing iron overload classifications between the initial and repeat acquisitions the assigned iron overload categories were the same for SENSE and all Compressed SENSE factors (see Table 18 and Table 19). Figure 27 shows the scatter and Bland-Altman plots for the repeated acquisitions. Table 21 gives the Spearman's rho and ICCs for the comparisons, all of which indicate a strong and significant correlation and agreement between repetitions. Figure 27 shows that the bias was largest between repetitions for Compressed SENSE 2 and Compressed SENSE 6, both of which had a positive bias for the second repeat of greater than 3.57 s^{-1}. The 95% LoA were greater than 7.14 s^{-1} across all acquisitions and were particularly wide for Compressed SENSE 4. The large bias and/or LoA were primarily driven by large differences between repetitions for the four vials with the highest R_2*. For the 7 vials with lower R_2* generally the difference between the initial measurement and the repeat was less than 3.57 s^{-1}, although for Compressed SENSE 4 there was one vial over this threshold and for Compressed SENSE 6 there were 3 vials over this threshold (although still below the 7.14 s^{-1} threshold).

	SENSE Initial			SENSE Repeat			Compressed SENSE 2 Initial			Compressed SENSE 2 Repeat		
Vial	mDIXON Quant R_2* (s^{-1})	mDIXON Quant R_2* converted to LIC (mg/g dw)	Iron overload category	mDIXON Quant R_2* (s^{-1})	mDIXON Quant R_2* converted to LIC (mg/g dw)	Iron overload category	mDIXON Quant R_2* (s^{-1})	mDIXON Quant R_2* converted to LIC (mg/g dw)	Iron overload category	mDIXON Quant R_2* (s^{-1})	mDIXON Quant R_2* converted to LIC (mg/g dw)	Iron overload category
2	25.17	0.25	Normal	28.05	0.33	Normal	24.66	0.24	Normal	23.23	0.20	Normal
3	61.145	1.26	Normal	61.91	1.28	Normal	60.72	1.25	Normal	61.88	1.28	Normal
4	97.426	2.27	Insignificant	96.45	2.25	Insignificant	98.29	2.30	Insignificant	95.42	2.22	Insignificant
5	117.79	2.84	Insignificant	117.08	2.82	Insignificant	115.99	2.79	Insignificant	115.35	2.77	Insignificant
6	166.17	4.20	Mild	167.97	4.25	Mild	161.10	4.06	Mild	162.12	4.08	Mild
7	210.26	5.43	Mild	208.05	5.37	Mild	211.91	5.48	Mild	210.87	5.45	Mild
8	316.66	8.41	Moderate	314.41	8.35	Moderate	317.58	8.44	Moderate	319.30	8.49	Moderate
9	425.1	11.45	Moderate	429.46	11.57	Moderate	437.93	11.81	Moderate	432.46	11.65	Moderate
10	635.6	17.34	Severe	637.16	17.39	Severe	658.41	17.98	Severe	664.08	18.14	Severe
11	796.88	21.86	Severe	780.59	21.40	Severe	839.39	23.05	Severe	856.42	23.52	Severe
12	813.71	22.33	Severe	797.62	21.88	Severe	924.20	25.42	Severe	955.68	26.30	Severe

Table 18 R_2* from SENSE and Compressed SENSE 2, converted to LIC using Hankins' equation[58] and categorised in terms of iron overload[56]

	Compressed SENSE 4 Initial			Compressed SENSE 4 Repeat			Compressed SENSE 6 Initial			Compressed SENSE 6 Repeat		
Vial	mDIXON Quant R_2* (s^{-1})	mDIXON Quant R_2* converted to LIC (mg/g dw)	Iron overload category	mDIXON Quant R_2* (s^{-1})	mDIXON Quant R_2* converted to LIC (mg/g dw)	Iron overload category	mDIXON Quant R_2* (s^{-1})	mDIXON Quant R_2* converted to LIC (mg/g dw)	Iron overload category	mDIXON Quant R_2* (s^{-1})	mDIXON Quant R_2* converted to LIC (mg/g dw)	Iron overload category
2	23.61	0.21	Normal	25.46	0.26	Normal	25.847	0.27	Normal	26.76	0.29	Normal
3	62.37	1.29	Normal	63.23	1.32	Normal	67.019	1.42	Normal	65.17	1.37	Normal
4	103.38	2.44	Insignificant	101.14	2.38	Insignificant	94.217	2.18	Insignificant	98.00	2.29	Insignificant
5	119.84	2.90	Insignificant	120.24	2.91	Insignificant	113.23	2.72	Insignificant	116.67	2.81	Insignificant
6	168.52	4.26	Mild	170.28	4.31	Mild	167.91	4.25	Mild	172.01	4.36	Mild
7	214.87	5.56	Mild	213.71	5.53	Mild	207.33	5.35	Mild	210.69	5.44	Mild
8	321.67	8.55	Moderate	317.48	8.43	Moderate	313.49	8.32	Moderate	318.33	8.46	Moderate
9	423.50	11.40	Moderate	426.32	11.48	Moderate	410.37	11.04	Moderate	417.46	11.23	Moderate
10	625.95	17.07	Severe	641.78	17.51	Severe	605.02	16.49	Severe	601.07	16.37	Severe
11	729.52	19.97	Severe	790.81	21.69	Severe	709.14	19.40	Severe	698.40	19.10	Severe
12	840.06	23.07	Severe	773.99	21.22	Severe	715.93	19.59	Severe	750.70	20.56	Severe

Table 19 R_2* from Compressed SENSE 4 and Compressed SENSE 6, converted to LIC using Hankins' equation[58] and categorised in terms of iron overload[56]

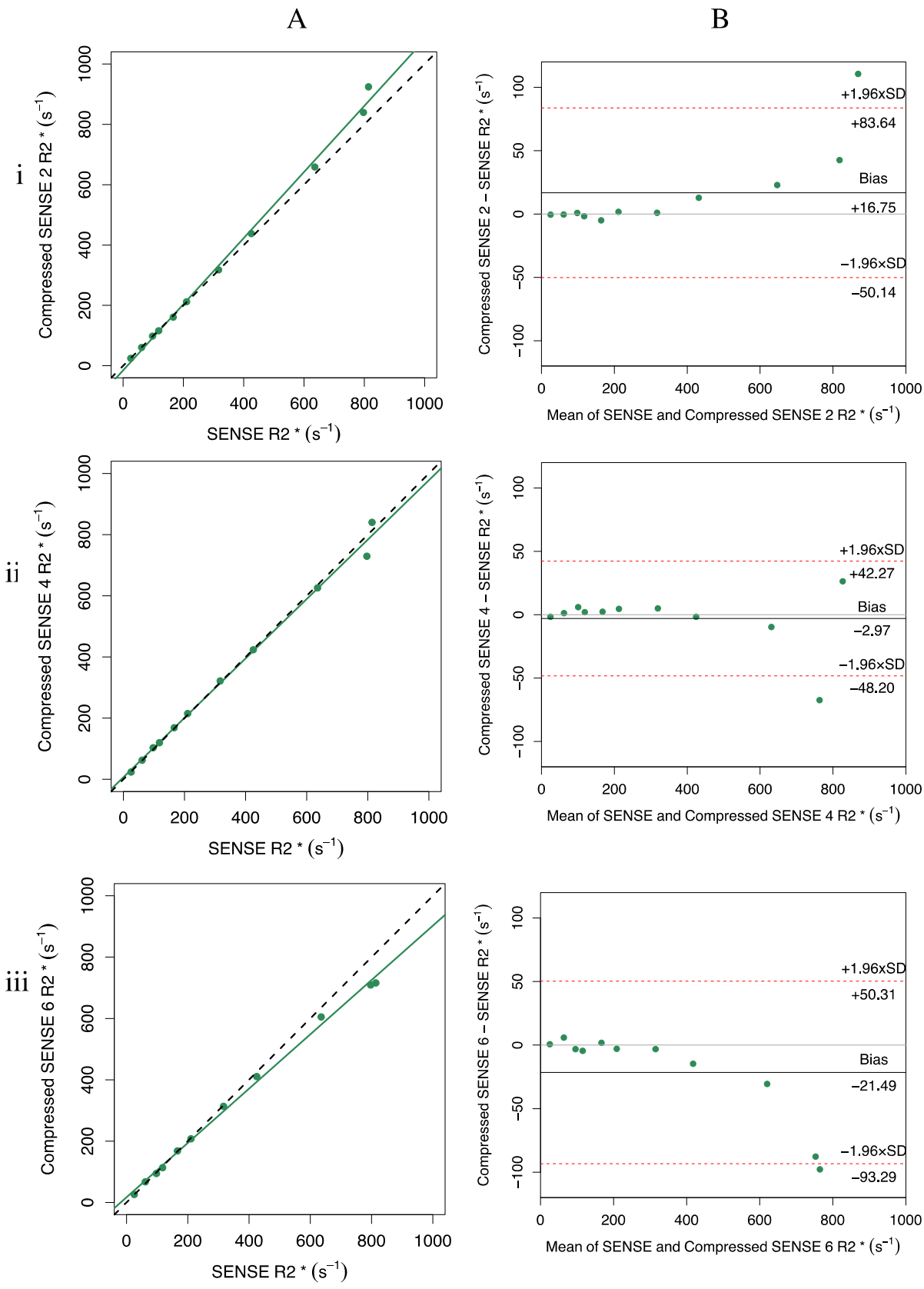


Figure 25 Scatter plots (A) and Bland Altman plots (B) comparing R2* results from Calimetrix phantom with SENSE and (i) Compressed SENSE 2, (ii) Compressed SENSE 4 and (iii) Compressed SENSE 6. For the scatter plots the dashed black line is the identity line and the solid green line is the line of best fit.

	SENSE	
	Spearman's rho	**ICC**
Compressed SENSE 2	1.00 (p <0.001)	0.99 (p <0.001)
Compressed SENSE 4	1.00 (p <0.001)	>0.99 (p <0.001)
Compressed SENSE 6	1.00 (p <0.001)	0.99 (p <0.001)

Table 20 Spearman's rho and Intraclass Correlation Coefficients for R_2* for comparison of SENSE and Compressed SENSE

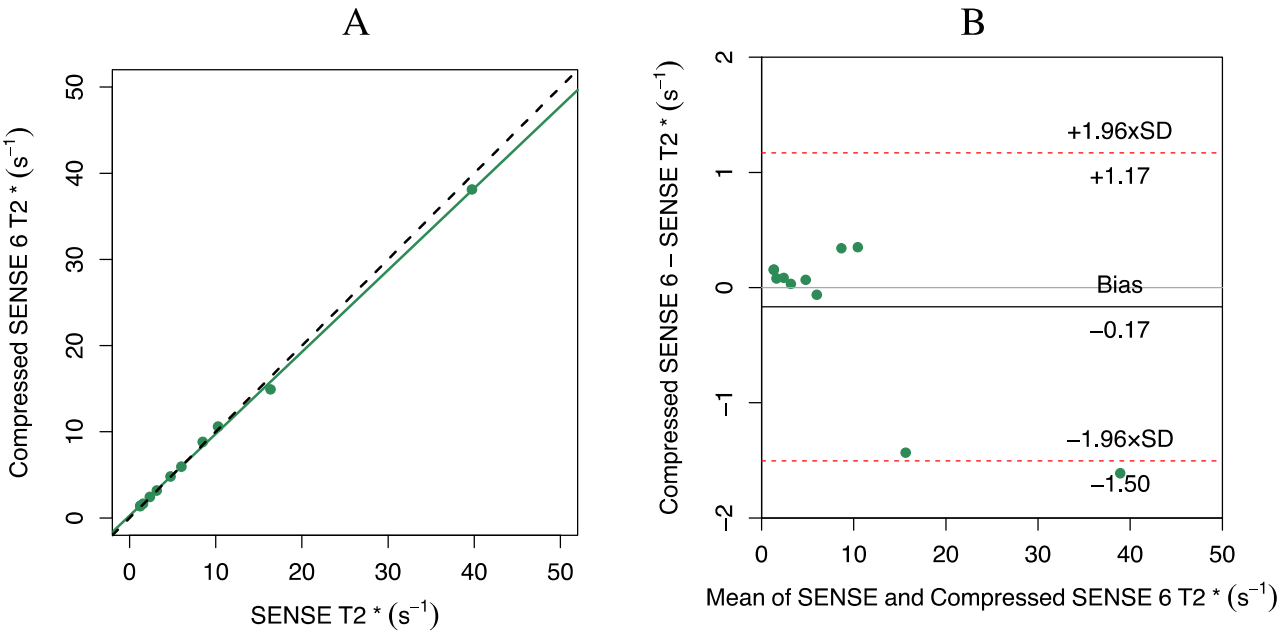


Figure 26 Scatter plot (A) and Bland Altman plots(B) comparing T_2* results from Calimetrix phantom with SENSE and Compressed SENSE 6. For the scatter plot the dashed black line is the identity line and the solid green line is the line of best fit.

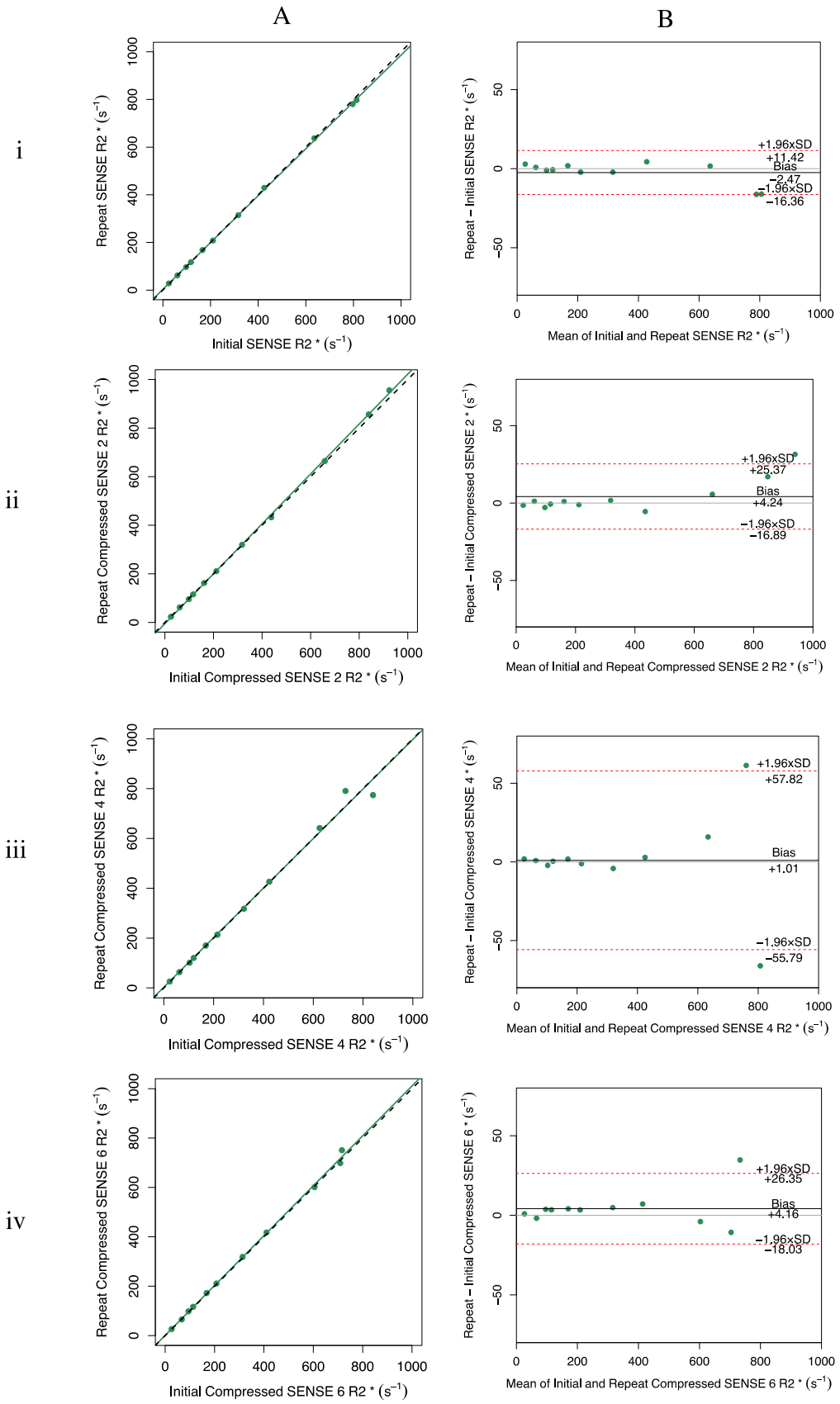


Figure 27 Scatter plots (A) and Bland Altman plots (B) comparing R2* measurements from repetitions of mDixon Quant scan on Calimetrix phantom using (i) SENSE, (ii) Compressed SENSE 2, (iii) Compressed SENSE 4 and (iv) Compressed SENSE 6. For the scatter plots the dashed black line is the identity line and the solid green line is the line of best fit.

	Spearman's rho	ICC
SENSE	1.00 (p <0.001)	>0.99 (p>0.001)
Compressed SENSE 2	1.00 (p <0.001)	>0.99 (p>0.001)
Compressed SENSE 4	0.99 (p <0.001)	>0.99 (p>0.001)
Compressed SENSE 6	1.00 (p <0.001)	>0.99 (p>0.001)

Table 21 Spearman's rho and Intraclass Correlation Coefficients for comparisons between repetitions of mDixon Quant sequence with difference acceleration factors

5.3.4. Discussion

In this study I compared R_2* and T_2* results from a phantom scanned with a 3D ME GRE sequence with and without compressed sensing (without compressed sensing: SENSE acquisition, with compressed sensing: Compressed SENSE acquisition), to assess if compressed sensing can be used to accelerate 3D ME GRE sequences for R_2* quantification. Given that the iron overload classification remained the same with and without compressed sensing, this indicates that the sequence with compressed sensing could probably be used for iron overload diagnosis. The large differences in R_2* (> 7.14 s^{-1}, equivalent to >0.2 mg/g dw difference in iron concentration) between the results with and without compressed sensing for some vials across all three Compressed SENSE factors indicate that it might not be appropriate to use compressed sensing for treatment monitoring, particularly if the patient has previously been scanned without compressed sensing: differences detected in R_2* between timepoints may not be due to physiological differences but due to the different method of acceleration. However, this is only for the vials corresponding to moderate/severe iron overload: the difference is less for lower R_2* vials, so if the patient has low iron concentration, it may be acceptable to use compressed sensing. The large differences also mean that it might not be appropriate to use the sequences for detecting iron sparing. However, given that the differences between the results with and without compressed sensing are much less for the vials with lower R_2*

(i.e., lower iron concentration, as we would expect in iron sparing), then it may be acceptable to use compressed sensing in this scenario.

The reason for the bias compared to SENSE could be due to a variety of reasons. It may be because the subsampling of k space is random for compressed sensing and not for SENSE, and this random subsampling may not adequately capture the higher frequency changes associated with the vials. The denoising may also blur the vials with the surroundings. Another reason may be that the random subsampling results in a different k-space sampling for each echo, and so the fitting of the T_2* decay curve may sometimes be using multiple points of estimated data with non-trivial errors. It is interesting that the bias was positive for the lowest compressed sensing factor, but negative for higher compressed sensing factors. It is unclear what the reason for this might be. The reason for the larger difference between SENSE and Compressed SENSE values for higher R_2* is probably due to difficultly of fitting the T_2* curve when the T_2* decay is fast, especially given that the first TE was 1.24 ms, by which the signal will already have dropped by a large amount. This means that any issues arising from the Compressed SENSE method are likely to be amplified.

In contrast to R_2*, the bias and LoA for T_2* were small. Boyarko *et al.* compared T_2* with and without compressed sensing in 36 children and young adults[87]. They found only a moderate correlation and a larger bias (-3.0 ms) and wider LoA (+ 6.0 ms and -12.0 ms) compared to our study. They suggested that this bias may be due to motion between their calibration scan for compressed SENSE and the actual scan, as the calibration scan used was from the SENSE scan. In the phantom there was no motion, and hence this could be the reason for the stronger correlation and improved bias. We found our results indicate that, although the T_2* results look promising, clinically, it might not be acceptable to use compressed sensing in some clinical scenarios. Additionally, the differences between

Compressed SENSE and SENSE were greatest for longest T_2*, i.e., low R_2*, but as shown by the our results, the lowest R_2* results have the smallest difference in R_2* between SENSE and Compressed SENSE and thus the smallest differences in iron.

Compressed sensing generally had poorer results in terms of repeatability compared to the acquisition without compressed sensing, with larger differences between the initial and repeated acquisitions, again indicating that it might not be suitable for use in patient monitoring for treatment or detection of cancer, as real physiological differences will be difficult to identify. The iron overload classifications were still the same, indicating it may be appropriate to use in that scenario. Once again, the differences between repeats were a lot less for lower R_2* vials, so it may depend on the R_2* of the patient as to whether it is acceptable to use compressed sensing or not. The reason for the variability between repeats could be due to the unpredictable, random subsampling of k space that compressed sensing involves. SENSE results did also show some variability between repeats, although generally it had a smaller bias and LoA compared to the Compressed SENSE sequences. This indicates that even without compressed sensing, care should be taken when using these sequences for monitoring. The reason for the differences between the repeats could be due to difficulties of estimating T_2* with 6 echoes, particularly for short T_2* (i.e., higher R_2*), where the decay is rapid. This means estimates are likely to be quite different. Using compressed sensing may amplify this. Therefore, a solution may be to optimise the sequence for the application, e.g., short T_2*, and then the repeatability may improve.

Aside from the study by Boyarko *et al.*[87] there is limited literature on the impact of compressed sensing on T_2* or R_2* results. There have been studies investigating use of compressed sensing in various non-quantitative clinical applications. For example, one study looking at ten different clinical applications found that, in general, compressed

sensing had no statistically significant impact on image quality[50]. However in other studies artifacts unique to compressed sensing were found[52,88], such as global ringing and wax-layer artifact. One quantitative study using compressed sensing performed by Berg *et al.*[89] contained five volunteers who had their brains scanned with a modified variable flip angle radiofrequency (RF) spoiled ME gradient-echo, from which R_2* can be estimated. R_2* was estimated from different regions in the brain for each of the volunteers. The difference in mean R_2* of non-accelerated compared to accelerated was small (e.g., for no acceleration compared to compressed sensing factor 6 the bias was -0.42 s^{-1}). The smaller bias compared to our study could be due to the low values of R_2* being measured (less than 26 s^{-2}). For lower R_2*, we also found much smaller differences. Another difference between the studies is that their study was performed at 3 T. It was on a Philips scanner, like our study. The lack of studies indicates a need for further work in this area.

A limitation of this study is that it used a phantom, and some of the differences between results with and without compressed sensing, and for repeatability, may be different *in vivo*, where the structure of the liver is quite different in comparison to the phantom. However, the phantom has the advantage that it can be scanned multiple times and it has a range of R_2* values, which would not be possible to assess in a healthy volunteer population. Additionally, it is not impacted by artifacts such as motion, and therefore can be used to assess differences in R_2* with fewer confounds. It is not envisaged that this work would be used in isolation but gives a useful foundation for future studies. It would be useful to use a phantom which was more representative of the liver, for example using iron instead of manganese chloride to modulate R_2* and also using fat. These types of phantoms are currently in development by some groups[84,86], and therefore, at a later date, it would be useful to repeat this study on such a phantom.

Another limitation of this study is that it only covers one field strength and vendor. The results may be different at 3 T and for different vendor sequences.

5.4. Summary

The first phantom study showed that using different sequences yields different R_2* values. Although these differences may have little impact on categorising iron overload in patients, they could have impact if a patient is scanned with different sequences over time. Additionally, it highlighted the requirement to optimise parameters for the particular application, whether that be for high iron, or for detecting iron sparing. Finally, it showed some of the limitations of processing methods when using a phantom, which MR operators should consider.

The second phantom study showed that using compressed sensing can yield different R_2* results to data acquired without compressed sensing. However, these differences may not influence iron overload classification and may be confined to cases with moderate or severe iron overload. Further evaluation is needed *in vivo*.

6. Initial Evaluation of R_2* sequences: *In vivo*

6.1. Introduction

Following on from the phantom work assessing 3D ME GRE sequences for R_2* estimation with and without compressed sensing (Chapter 5), the next step was to evaluate a 3D ME GRE sequence *in vivo*. In this chapter I assessed a slightly modified version of the Philips mDixon Quant sequence against the requirements for iron overload diagnosis, treatment monitoring and detection of iron sparing. This included assessment of sequences with and without compressed sensing, used to accelerate the acquisition.

I compared R_2* measurements from 3D acquisitions with/without compressed sensing to those from 2D acquisitions (the "ground truth"). In addition, I assessed the quality of R_2* maps throughout the liver, with the application of iron sparing primarily in mind.

Again, although the aim of this work was initially to evaluate these sequences for the application of detecting iron sparing, the applications for diagnosing iron overload and monitoring iron during therapy are also considered.

I helped set up the sequence on the scanner (at Advanced Medical Imaging Singapore) alongside Sue Francis from Nottingham and members of the Perspectum Imaging Applications team. I set up the processing pipeline for the data including producing the code to help with this. I then oversaw the analysis of the data by a graduate intern at Perspectum, as well as the statistical analysis. I evaluated the results and formulated the conclusions.

6.2. Method

6.2.1. Patients

Patients were recruited as part of the ELEGANCE Study, an investigator-initiated multi-centre, longitudinal cohort study of patients in Singapore at increased risk of HCC (ClinicalTrials.gov: NCT04965259) under the auspices of the Asia-Pacific Hepatocellular Carcinoma (AHCC) Trials Group. All patients gave their written, informed consent.

I report results for 81 patients. Every patient had chronic liver disease: 29 patients had cirrhosis (with different underlying aetiologies), 43 had non-cirrhotic Hepatitis B and 9 had Non-Alcoholic Fatty Liver Disease (NAFLD)/Non-Alcoholic Steatohepatitis (NASH).

6.2.2. MRI Acquisition

Patients were scanned on a 3 T Ingenia Philips scanner at Advanced Medical Imaging, Singapore with 2D and 3D R_2* mapping sequences.

The 2D sequence used was MOST, for further details see Section 3.7. The 3D ME GRE sequence was based on mDIXON Quant. However, some modifications were made to the "default" vendor parameters, most notably increasing the number of echoes from 6 to 12, to yield a sequence that is better able to quantify longer T_2* associated with the lower iron levels in nodules of HCC[17]. The SENSE parallel imaging technique was used to accelerate the acquisition, but the breath-hold length was long for this sequence, around 19.0 s (the number of slices varied between patients so there was some difference in breath-hold length between patients). I refer to this sequence as mDIXON SENSE. A version of mDIXON Quant was made using compressed sensing (Compressed SENSE factor 4.5), which had a breath-hold length of around 10 s (depending on the number of slices), a reduction in time by approximately a half. I refer to this as mDIXON

Compressed SENSE. For acquisition parameters for the mDixon Quant sequence see Table 22, MOST parameters are given in Section 3.7. 2D ME GRE and 3D ME GRE without Compressed SENSE was acquired for all patients, while 76 of the 81 patients also had 3D ME GRE with Compressed SENSE.

Parameter	mDIXON Quant	mDIXON Quant with Compressed SENSE
Field of View (mm x mm)	400 x 350	400 x 350
Acquisition Matrix size	160 x 140	160 x 140
Acquired Resolution (mm x mm)	2.50 x 2.50	2.50 x 2.50
Reconstructed voxel size (mm x mm)	2.08 x 2.08	2.08 x 2.08
Acquired slice thickness (mm)	6.00	6.00
Reconstructed slice thickness (mm)	3.00	3.00
Number of slices per slab	~55 (adjusted according to liver size)	~55 (adjusted according to liver size)
Number of slabs	1	1
Number of repeats	None	None
TR (ms)	10.24	10.24
Number of Echoes	12	12
Initial TE (ms)	1.00	1.00
Final TE (ms)	9.09	9.09
Echo spacing (ms)	0.74	0.74
Bandwidth (Hz/Pixel)	2232	2232
Flip Angle	3	3
Parallel Imaging: SENSE Factor	2 (Anterior Posterior direction)	N/A
Compressed SENSE Factor	N/A	4.5
Partial Fourier	Off	Off
Acquisition duration (s)	19.0 s for 55 slices	10.1 s for 55 slices

Table 22 Sequence parameters

6.2.3. Analysis and Statistics

MAGO[65] was used to generate T_2* and R_2* maps from the 2D and 3D ME GRE sequences. Analysis was performed by a trained analyst, the graduate intern who I was supervising, with several months experience. All analysis was reviewed by me (a senior analyst with 3.5 years' experience). Although T_2* maps were used for the image analysis, all statistics were performed using R_2*.

ROI analysis

Firstly, to verify that 3D acquisitions can be used in place of 2D acquisitions, ROI analysis was performed on the 2D ME GRE T_2* map (located in the centre of the liver) and on the matching slice of the 3D acquisitions (located using slice location and visual assessment). Three 15 mm diameter circular ROIs were placed, avoiding major vessels and artifacts. Median R_2* for pixels contained within the pooled ROIs was calculated. The ROI median R_2*s were compared using scatter plots, Spearman's rho (correlation), (ICC, to assess agreement) and Bland Altman plots (bias and LoA).

Segmentation

For the quality assessment throughout the liver, the liver was segmented in multiple slices of the 3D acquisitions. Segmentation was performed semi automatically, with the liver firstly being automatically segmented using a U-NET segmentation method [90], followed by manual adjustments. Artifacts were not excluded from the segmentation. Vessels were excluded using hysteresis thresholding[91] with the thresholds based on manual segmentation of vessels in 10 cases. Segmentation was performed on the central slice, as well as two upper and lower slices (the peripheral slice at the top and bottom of the liver was identified, and slices selected for segmentation 15 mm and 30 mm from that slice).

Each case was assessed for physiological heterogeneity, as the subsequent analysis depended on the assumption that all major heterogeneity in R_2* was the result of artifacts and was not physiological. Although heterogeneity can occur in liver iron concentration, it generally happens in patients with iron accumulation (unlike this population), and even then it is not often seen on MRI[92]. Additionally, the senior analyst confirmed that for all cases any large differences in R_2* were the result of artifacts, given their characteristics,

for example being positioned near the lungs/heart, in areas of gradients in the field maps (hence indicating field heterogeneity causing T_2* shortening).

Once segmentation had been performed, quality within the liver was assessed. One method used was to score the central slice for artifacts, giving scores for both extent and magnitude of the artifact (scored 0-3, 0 being best). The senior analyst (me) performed the scoring. To aid with detecting artifacts, thresholding of R_2* was used to identify areas where R_2* was higher (i.e., shorter T_2*), as T_2* shortening is a typical indicator of artifacts such as field heterogeneity and motion. The thresholds were chosen to reflect 0.1 mg/g dw (R_2* 7.3 s^{-1}), 0.2 m/g (R_2* 14.7 s^{-1}) and 0.3 mg/g dw (R_2* 22.0 s^{-1}) increases in LIC (see details of the conversion in Section 6.2.4). The analyst made a note of the potential causes of the artifact. Noise was not counted as an artifact, so diffuse distribution of higher R_2* pixels was not scored as an artifact. The scoring system is shown in Table 23. Tables of the scores were produced to assess the frequency of artifacts, both in terms of extent and magnitude.

Score	Extent	Magnitude
0	No artifact.	No artifact
1 - Mild	<12.5% of liver	Largest proportion of pixels in impacted area are ≥7.3 s^{-1} but <14.7 s^{-1} from ROI median.
2 - Moderate	12.5%-50% of liver	Largest proportion in impacted area ≥14.7 s^{-1} from ROI median, but generally < 22.0 s^{-1}
3 - Severe	More than 50% of liver	Largest proportion in impacted area ≥14.7 s^{-1} from ROI median, and generally ≥22.0 s^{-1} from ROI median

Table 23 Scoring system. 7.3 s^{-1} is equivalent to 0.1 mg/g dw difference in LIC, 14.7 s^{-1} is equivalent to 0.2 mg/g dw LIC and 22.0 s^{-1} is equivalent to 0.3 mg/g dw LIC, using Alam and Hankins' equations.

Within-slice heterogeneity was assessed using the R_2* interquartile range (IQR) of the segmentations for all five segmented slices. A slice with a large IQR indicates more heterogeneity and so is likely to be of poor quality. IQRs of the segmentation R_2* were plotted in box plots.

Quality between slices was also assessed. The median R_2* of the segmentation of the upper and lower slices were compared with the median R_2* of the central slice. Assuming any heterogeneity is due to artifacts, a good quality volume would result in comparable R_2* throughout all the slices. Comparisons were performed using scatter plots, Spearman's rho (correlation), ICC (agreement) and Bland Altman plots (bias and LoA).

Another way of assessing quality is to use the R^2 goodness-of-fit maps. This gives the proportion of variance in the signal explained by the model and so provides a measure of fit: how close the predicted signal values of the model (which is used to generate T_2*) are to the actual signal values. If there are deviations in the signal away from the predicted values this indicates there may be problems with the quality i.e., an artifact might be causing deviation away from the predicted T_2* decay curve. R^2 is given as a fraction: if the R^2 is 1 then the model values identically match the actual values. R^2 were generated and the median R^2 calculated for the segmentations. The results for the different slices and acquisitions were presented as box plots.

6.2.4. Assessment against clinically important thresholds

To assess whether differences in R_2* between acquisitions and within the liver are clinically significant, R_2* differences need to be converted to iron concentration. Therefore the 3 T R_2* was firstly converted to 1.5 T R_2* using the conversion reported by Alam *et al.*[60] and then converted to LIC using Hankins' equation[58].

Clinically acceptable differences in iron concentration due to non-physiological variation in R_2* vary with application and context. The same thresholds, as outlined in Chapter 4 and 5 can also be applied to this context:

- Iron overload classification: R2* differences equivalent to up to 0.5 mg/g dw, i.e., at 3 T differences of up to 36.6 s^{-1}.

- Treatment monitoring e.g., of iron chelation: R2* differences equivalent to up to 0.2 mg/g dw, i.e., at 3T differences of up to 14.7 s^{-1}.
- Detection of iron sparing: R2* differences equivalent to up to 0.1 mg/g dw, i.e., at 3 T differences of up to 7.3 s^{-1}

These thresholds apply to between sequence variability: the differences between sequences should be less than the thresholds to give confidence in the sequences (particularly the 3D ME GRE sequences compared to the robust 2D ME GRE sequence), as well as then being acceptable to use the sequences interchangeably in treatment monitoring, for example. They also apply to variability within the liver due to artifacts: if the non-physiological variability within the liver is less than these thresholds then they meet the requirements for the application.

6.2.5. Statistical Tests

As in previous sections, statistical tests included Spearman' rho (correlation), ICC (agreement) and Bland Altman plots (bias and LoA). p values of < 0.05 were classed as statistically significant.

6.3. Results

6.3.1. ROI analysis

Figure 28 shows scatter plots and Bland Altman plots showing the median of R_2* from the ROIs compared between acquisitions. For all comparisons, Spearman's rho was $\geq$ 0.90 (significant) and ICC was $\geq$ 0.98 (significant). The points lie close to the identity line and the bias was $< 1s^{-1}$ in magnitude in all cases. Additionally, the upper and lower LoA were less than $10s^{-1}$ in magnitude. Therefore, both the magnitude of the bias and the LoA lie within the threshold of variability acceptable for diagnosis of iron overload. They also lie within the threshold for using the sequences interchangeably for iron monitoring

and within/close to the thresholds for iron sparing detection. Full details of the statistical test results are given in Table 24.

6.3.2. Scoring

The results of the scoring are shown in Table 25 and Table 26, which show that 38% and 49% of cases were impacted by artifacts for SENSE and compressed SENSE acquisitions respectively. Neither acquisition resulted in any cases with a score of 3 for extent. For SENSE there were no cases where both the extent was scored 2 or 3 and the magnitude was scored 2 or 3, but there were 6.6% cases where this was the case for the Compressed SENSE version. The analyst reported that all artifacts were due to motion and/or field heterogeneity and particularly impacted the dorsal area of the liver (due to proximity to the lungs) and the left lobe (due to proximity to the heart).

6.3.3. Within slice heterogeneity: Inter-quartile range

Figure 29 shows box plots of the IQR of the R_2* from the segmentation of the different slices. The IQR increased the further up the liver the slice was located. For all slices the majority of cases had IQR below 36.6 s^{-1}. This means that the majority are within the threshold of variability acceptable for diagnosis. Additionally, the median IQR across patients was below 14.7 s^{-1} for all slices apart from the dome of the liver, and so the median IQR lay within the threshold of variability acceptable for diagnosis of iron overload and monitoring of treatment such as iron chelation. However, the majority of cases had IQR above 7.3 s^{-1}, the threshold of variability in relation to iron sparing, which is discussed further in the discussion. The results for SENSE and Compressed SENSE acquisitions are similar.

6.3.4. Between slice heterogeneity

Between slice heterogeneity was assessed by comparing the median of R_2* from the segmentation of the central slice to those of the outer slices. Examples of the results

plotted in scatter plots and Bland Altman plots are shown in Figure 30, where the central slice was compared to the base slice. Appendix B contains the plots for comparisons of the central slice with the inferior, superior and dome slices. The results were essentially similar between SENSE and Compressed SENSE. Spearman's rho was significant for all comparisons but was lowest for the central slice vs dome comparisons. ICC was ≥ 0.90 (significant). Full details of Spearman's rho and ICCs can be found in Table 27 and Table 28.

The Bland Altman plots show that bias was negative in comparison with the central slice (i.e., R_2* lower) for the lower slices, but became increasing positive for the upper slices. The bias had a magnitude of less than 2 s^{-1} apart from for the comparison with the dome slice, where for SENSE the bias was still less than the 7.3 s^{-1} threshold (for iron sparing detection acceptable variability) but for Compressed SENSE the bias was 8.25 s^{-1}. Although some of the upper and lower LoA had magnitude greater than 7.3 s^{-1} they were generally 9 s^{-1} and less. For the dome the upper limit of agreement for both SENSE and Compressed SENSE was greater than 19 s^{-1}. Full details of the Bland Altman statistics are given in Table 27 and Table 28.

6.3.1. Assessment of fit

The box plot comparing the R^2 results between the different slices and acquisitions is shown in Figure 31. R^2 was generally above 0.9. There was a drop in R^2 in the outermost slices (although still mostly above 0.9). Compressed SENSE gave sightly poorer results than SENSE.

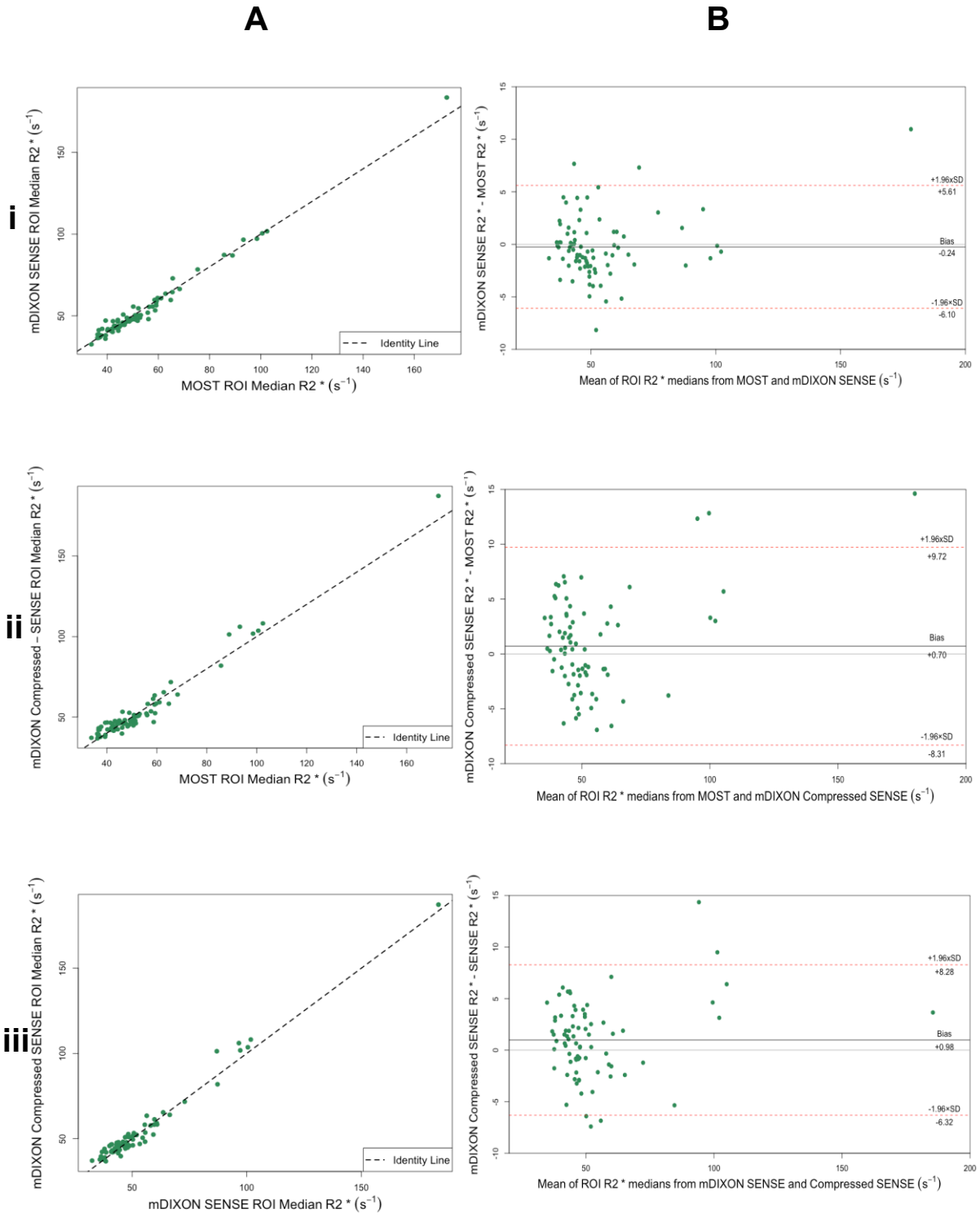


Figure 28 (A) Scatter plot (B) Bland Altman plots, comparing (i) 2D ME GRE (MOST) R_2* and 3D ME GRE with SENSE (mDIXON SENSE) (ii) 2D ME GRE (MOST) and 3D ME GRE with Compressed SENSE (mDIXON Compressed SENSE) (iii) mDIXON SENSE and mDIXON Compressed SENSE

	Spearman's rho	Spearman's rho p value	ICC	ICC p value	Bland Altman bias	Bland Altman lower limit of agreement	Bland Altman upper limit of agreement
MOST and SENSE	0.96	p <0.001	0.99	p <0.001	-0.24	-6.10	5.61
MOST and C-SENSE	0.90	p <0.001	0.98	p <0.001	0.70	-8.31	9.72
SENSE and C-SENSE	0.91	p <0.001	0.98	p <0.001	0.98	-6.32	8.28

Table 24 Statistical tests between R2* results from different acquisitions

		Extent			
		0	**1**	**2**	**3**
Magnitude	**0**	50 (61.7 %)	0	1 (1.2 %)	0
	1	0	16 (19.8 %)	4 (4.9 %)	0
	2	0	1 (1.2 %)	0	0
	3	0	9 (11.1 %)	0	0

Table 25 Scoring for mDIXON SENSE R_2* central slice

		Extent			
		0	**1**	**2**	**3**
Magnitude	**0**	39 (51.3 %)	0	1 (1.3 %)	0
	1	0	14 (18.4 %)	11 (14.5 %)	0
	2	0	0	1 (1.3 %)	0
	3	0	6 (7.9 %)	4 (5.3 %)	0

Table 26 Scoring for mDIXON Compressed SENSE R_2* central slice

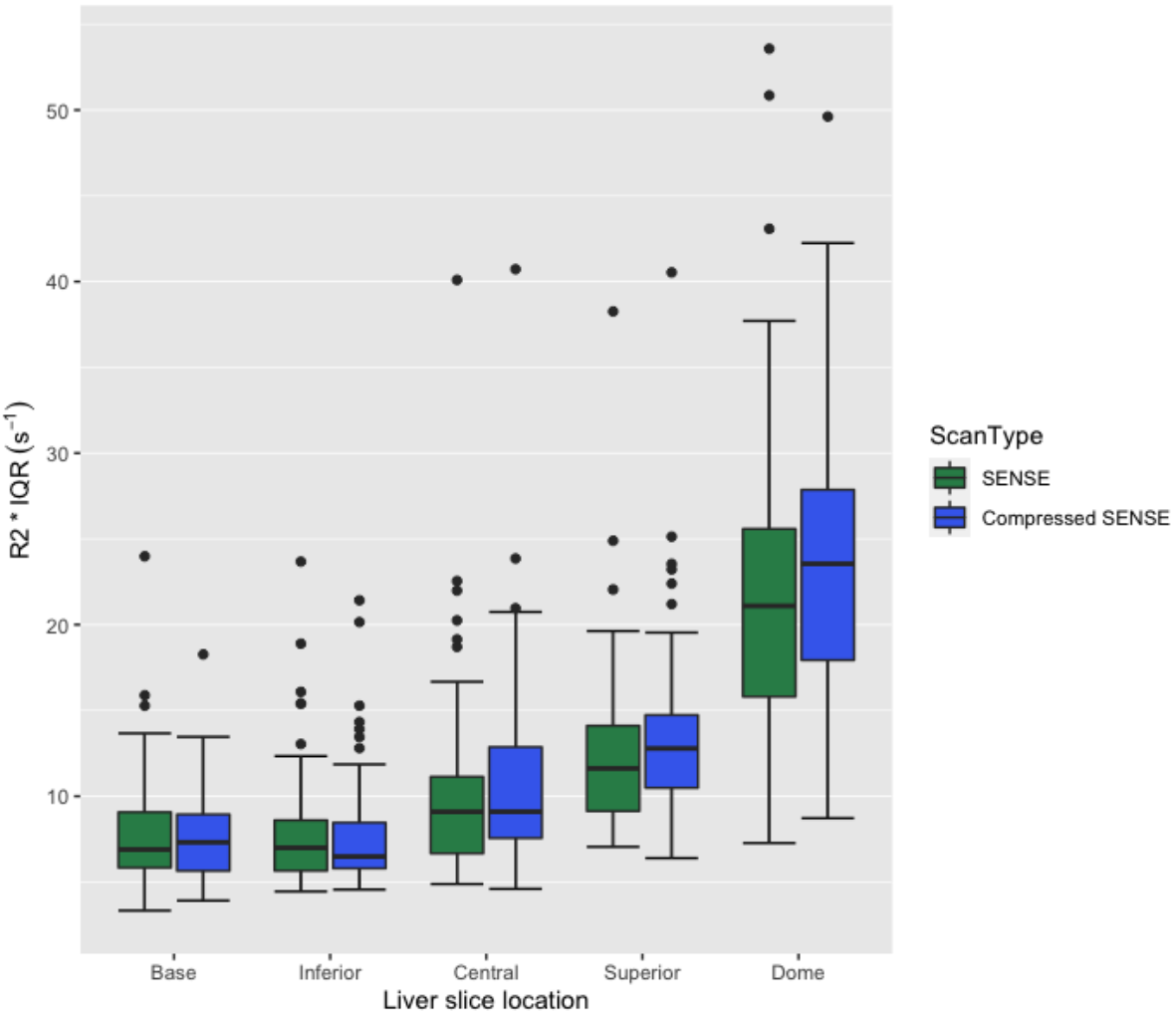


Figure 29 Box plot showing interquartile range of R_2* within segmentation at different slice locations for both SENSE and Compressed SENSE.

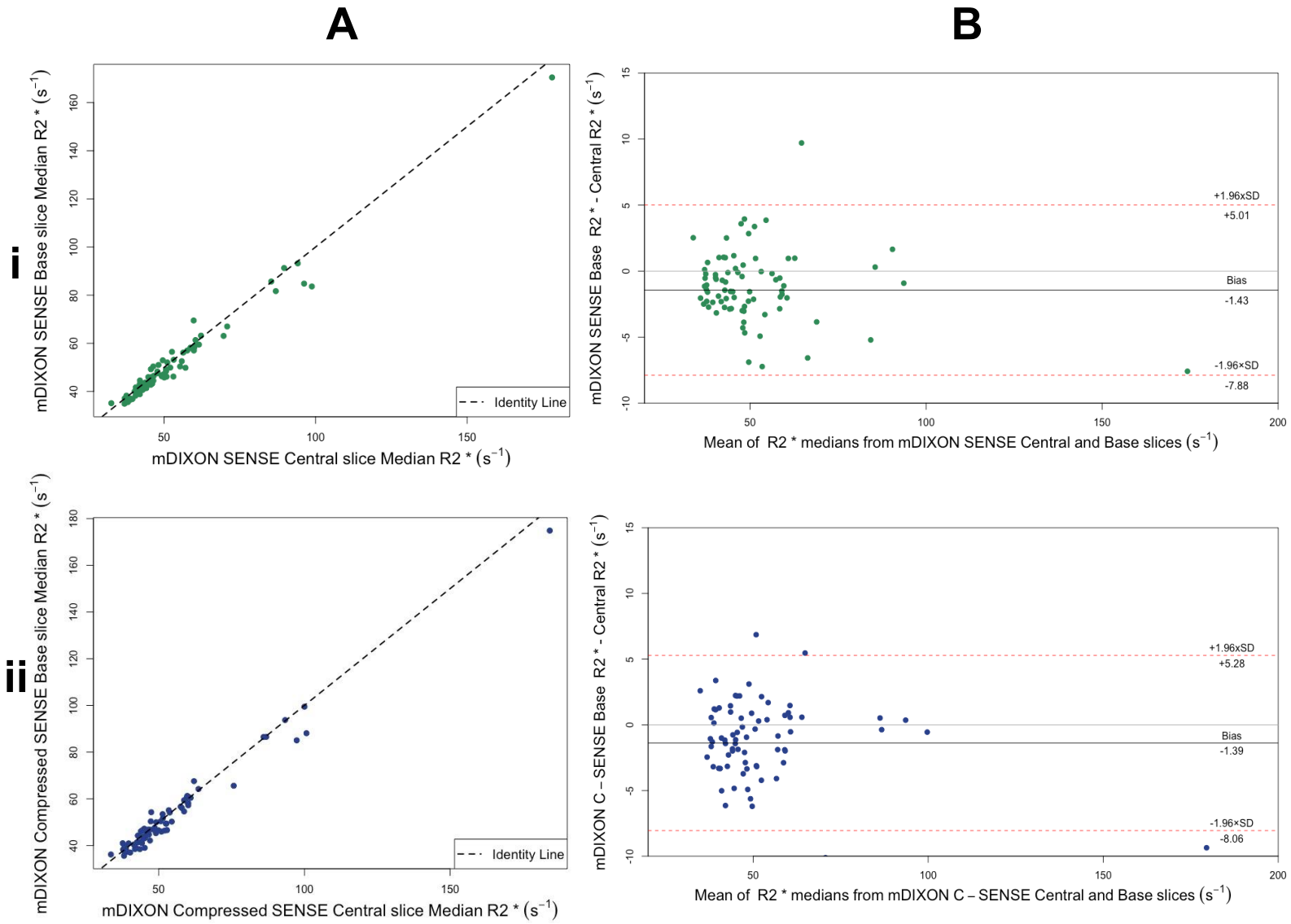


Figure 30 (A) Scatter plot (B) Bland Altman plots, comparing the R_2* median of the segmentations of the central slice and base slice (i) For mDIXON SENSE and (ii) mDIXON Compressed SENSE

	Spearman's rho	Spearman's rho p value	ICC	ICC p value	Bland Altman bias	Bland Altman lower limit of agreement	Bland Altman upper limit of agreement
Central slice + Base	0.97	p < 0.001	0.98 (0.97, 0.99)	p <0.001	-1.43	-7.88	5.01
Central slice + Inferior	0.98	p < 0.001	0.99 (0.97, 0.99)	p <0.001	-1.46	-6.80	3.87
Central slice + Superior	0.98	p < 0.001	0.99 (0.98, 1.00)	p <0.001	1.38	-2.31	5.06
Central slice + Dome	0.85	p < 0.001	0.91 (0.52, 0.97)	p <0.001	6.37	-6.34	19.07

Table 27 Statistical tests between R2* results from different slices for mDIXON SENSE

	Spearman's rho	Spearman's rho p value	ICC	ICC p value	Bland Altman bias	Bland Altman lower limit of agreement	Bland Altman upper limit of agreement
Central slice + Base	0.95	p < 0.001	0.98	p <0.001	-1.39	-8.06	5.28
Central slice + Inferior	0.96	p < 0.001	0.98	p <0.001	-1.76	-8.12	4.60
Central slice + Superior	0.93	p < 0.001	0.99	p <0.001	1.98	-4.17	8.13
Central slice + Dome	0.82	p < 0.001	0.90	0.01	8.25	-2.82	19.32

Table 28 Statistical tests between R2* results from different slices for mDIXON Compressed SENSE

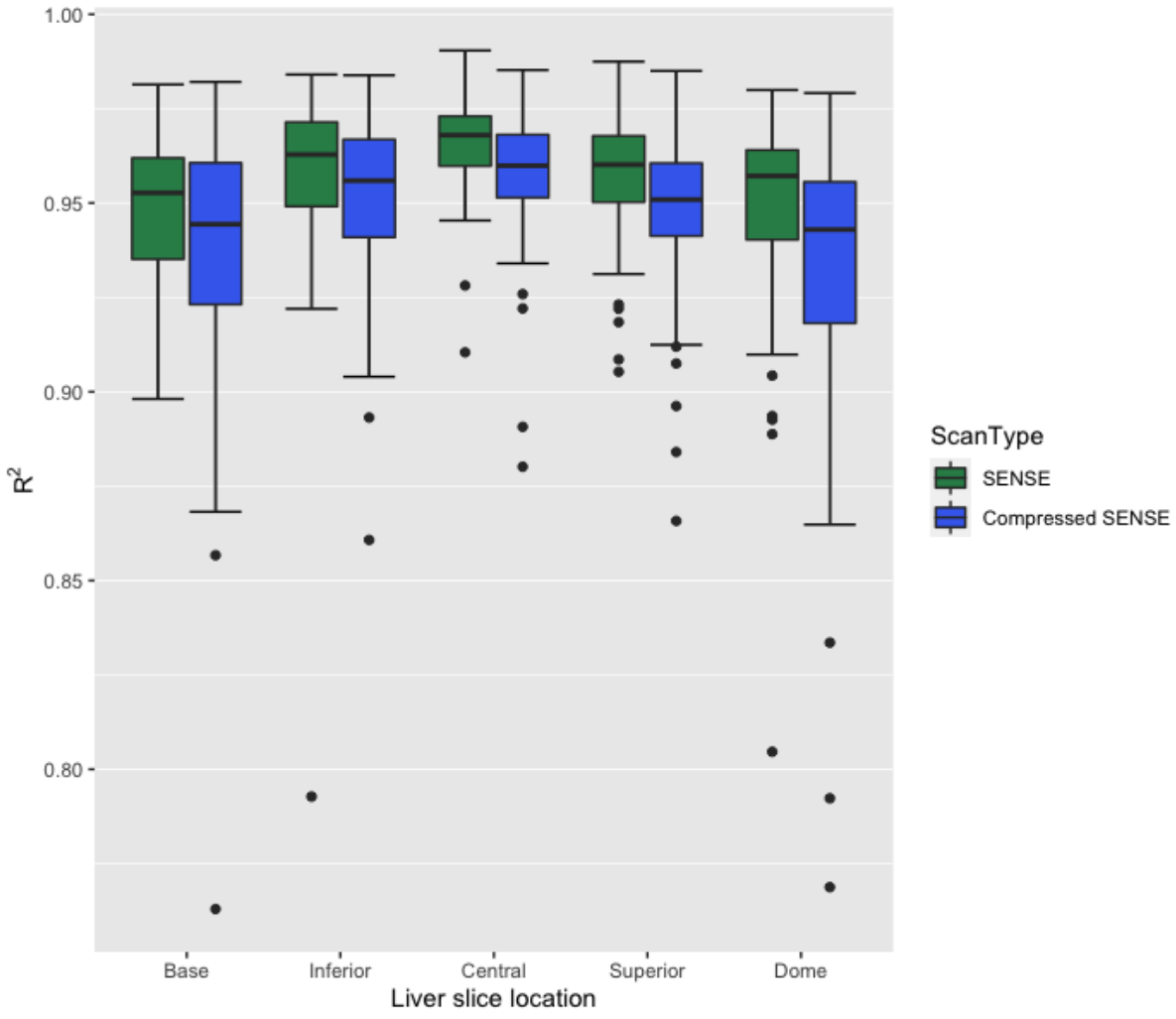


Figure 31 Box plot of Median R^2 of segmentations at different slice locations for both SENSE and Compressed SENSE.

6.4. Discussion

Given the increasing availability of 3D ME GRE sequences, it is important that the results are comparable to the 2D ME GRE sequences which have been used extensively. Additionally, given the potential utility of 3D acquisitions for identifying areas of heterogeneity in the liver, for example iron sparing as an indicator HCC, it is important that the R_2* values throughout the liver can be converted to accurate iron concentrations. In this study it was found that the results were comparable to 2D ME GRE sequences; but

that there was some non-physiological variability in R_2* values throughout the liver which may limit the use of these sequences for detection of physiological variability.

R_2* derived from ROIs in the 2D ME GRE sequence and in a single slice of the 3D ME GRE acquisitions were compared. Strong correlation and strong agreement were found between the 3D ME GRE sequences and the 2D ME GRE sequence. Importantly, the bias and LoA were such that they were lower than the thresholds of variability acceptable for detecting iron overload. Together, these findings suggest that it may be acceptable to use 3D sequences in place of 2D sequences. Additionally, the bias and LoA were lower than the thresholds for variability that were deemed acceptable for monitoring the impact of iron chelation treatment and close to/within the thresholds for iron sparing detection. These results show that the sequences can potentially be used interchangeably during monitoring.

When compared with the results in the phantom scan in Section 5.2, the difference between R2* measured using 3D ME GRE and the 2D ME GRE is less in this *in vivo* study. This may be due to the lower R2* in this patient population (despite being at 3 T, where we would expect higher R2*). Additionally, the confounding effects of the phantom, such as small vial size and phantom specific artifacts due to the processing not being optimised for phantoms, may have caused the difference in results.

Previous studies *in vivo* have found similar results. On a Siemens scanner at 1.5 T Henninger *et al*. compared 3D ME GRE acquisitions with a 2D acquisition and similarly found very small bias [81]. Similarly, Serai *et al.* investigated T_2* measurements on 2D and 3D ME GRE sequences at 1.5 T on a Philips scanner and found excellent agreement with small mean biases [82]. This work has extended this comparison to 3 T and a different patient population and has shown similar results.

Additionally, comparison was made between conventional parallel imaging (SENSE) and compressed sensing combined with parallel imaging (Compressed SENSE). Both SENSE and Compressed SENSE showed strong correlation and agreement and small bias both with the 2D ME GRE R_2* and with each other. This is in contrast with the results from the phantom in Section 5.3, where the bias was larger and LoA wider. However, the phantom included vials of much higher R_2*, and for the lower R_2* the differences were much less. When only vials of lower R_2* were included, the bias and LoA were more consistent with these *in vivo* findings (see Section 5.3.3).

The findings of the small bias and LoA is in contrast with a study by Boyarko *et al.* where they found only a moderate correlation between T_2* from 3D sequences with SENSE and compressed SENSE and suggested that Compressed SENSE cannot be used on 3D ME GRE sequences if accurate T_2* quantification is required [87]. The reason for the strong correlation and agreement found in this study may be due to acquiring a new calibration scan directly before the Compressed SENSE acquisition, instead of using the SENSE calibration scan (which is what Boyarko *et al.* did). Using the SENSE calibration scan can result in motion impacting the T_2* maps and hence the T_2* values [87].

However, using ROIs on a central slice does not make maximum use of the 3D acquisition. For example, detecting iron sparing requires assessment the entire volume. It was for this reason that we reported on quality within different slices of the liver. Note that this is in contrast with the above referenced papers. In Henninger *et al.*'s paper three ROIs 10-13 mm in diameter were used and mean R_2* compared [81]. In Serai's work they drew ROIs to encompass as much of the right lobe of the liver as possible (avoiding large vascular structures) on four slices in the mid-liver [82]. Similarly in Boyarko's study they drew ROIs on four slices in the mid-liver to cover as much of the liver as possible whilst

avoiding large vessels and artifacts [87]. In comparison, in this study, quality throughout the liver was assessed using a variety of methods.

By assessing the quality using scoring it was found that 38% and 49% of cases (for SENSE and Compressed SENSE acquisitions respectively) were impacted by artifacts, which is not ideal if being used to detect localised changes in iron. However, these were generally small in extent (less than 12.5 % of the liver) and/or small in magnitude (impact on R_2* corresponding to less than 0.2 mg/g dw), meaning that they would not significantly impact the use of the 3D ME GRE acquisitions for iron overload diagnosis and treatment monitoring.

Compressed SENSE had more cases with artifacts and artifacts were generally more severe in compressed SENSE compared with the SENSE acquisition. There are several potential explanations for this. One is that any artifacts in sub-sampled data may be extrapolated during the reconstruction process, as they will be the data that the algorithm is trying to optimise consistency with. Another is that for the SENSE acquisition the acceleration (and hence the subsampling) was only in the anterior posterior direction, whereas with Compressed SENSE the direction is not specified. Therefore, subsampling in the foot head direction may have increased the impact of through-plane artifacts. Additionally, a higher acceleration factor was used with Compressed SENSE compared to SENSE. It is important to note that using compressed sensing could hold potential risks if used in acquisitions for detection of iron sparing. The reason for this is that if the magnitude of iron sparing is small compared to the liver parenchyma, the sparsifying transform may result in this small difference becoming non-discernible on the reconstructed images. Additionally, if the tumour is small, the sampling may not be sufficient to delineate the tumour.

The IQRs of R_2* from the segmented liver across patients were largely within the variability threshold of detecting iron overload, indicating that sampling in different areas in the liver may be acceptable in that context. However, for detecting changes related to treatment such as chelation therapy more care is needed as there are many cases where the IQR is above the threshold of variability for this application. For iron sparing detection, most cases had IQR above the associated variability threshold. Caution is particularly needed in the highest slices, where the IQR was largest. The large IQRs, particularly for the higher slices, is probably due to the impact of motion and B_0 field heterogeneity due to the lungs and heart. This is consistent with work that found that B_0 inhomogeneities were greatest in segments of the liver close to the lung base [93], although they also found higher B_0 inhomogeneities in the segment near the transverse colon as well.

As regards comparisons between slices, R_2* was higher in the higher slices, as shown by the increasing positive bias in comparison to the central slice. However strong correlation and agreement was found between each of the slices with the central slice, with bias of less than or close to the threshold of variability for iron sparing detection. The LoA were slightly higher than the threshold for some slices, but only by a small amount apart from for the dome, where the upper limit of agreement was much higher. This indicates that using any slices for iron overload diagnosis is probably acceptable, but caution should be taken when using the dome slice, and detection of iron sparing in the upper slices of the liver may be challenging.

Increasing R_2* higher in the liver is consistent with being closer to the lungs and heart, and thus being more susceptible to field heterogeneity and motion. In a paper by Hernando *et al.* on R_2* mapping in the presence of B_0 field variations [94], an example is given of a patient scanned at 3 T where R_2* in the liver increased from 59 s^{-1} to 93 s^{-1} as

the slices were increasingly closer to the dome of the liver. In this study the same trend of increasing R_2* was found but the magnitude of the difference in R_2* across the liver was less. This may be due to the thicker slices in Hernando's study: their slice thickness was 8 mm[94].

Lastly, we assessed goodness-of-fit using R^2. Reassuringly, the R^2 values were generally above 0.9, although Compressed SENSE results were slightly worse than SENSE. This may be due to more subsampling in the Compressed SENSE acquisition compared to SENSE (the Compressed SENSE factor was higher than the SENSE factor).

The results for the various quality assessments, in particular the scoring and IQR results, suggest that there are limitations with the current 3D acquisitions, particularly for detecting iron sparing. These are primarily due to the impact of field heterogeneity and motion close to the lungs and heart. One change which might help improve the quality could be reducing the voxel size in the through-plane direction, which will reduce the impact of B_0 inhomogeneity near the lungs. Another potential method is to do B_0 corrections, which have been shown in improve consistency of R_2* throughout the liver[94]. However, this introduces another processing step, with extra potential for errors. Additionally, further investigation is needed into reducing artifacts on Compressed SENSE R_2* maps, given that artifacts were found more frequently, and there was greater variation within liver slices, compared to the SENSE acquisition. It may be that using a smaller compressed SENSE factor could improve the quality, whilst still providing some additional acceleration compared to SENSE.

One limitation of this study is that patients generally had low LIC, and the results may be different in a population with higher LIC. Therefore, for assessing the suitability for iron overload diagnosis and chelation treatment monitoring further studies are needed in such patient groups. Another limitation was that patients were scanned on one vendor

only. The vendor sequences are similar but further studies on other vendor scanners to confirm that the results are consistent with the results from this study. Another limitation is the full liver was not evaluated, only five slices. However, the five slices selected give a good sample of the whole liver as they were located both at the periphery of the liver and the centre.

6.5. Summary

The results indicate that the quality of acquisitions is such that it is acceptable to use 3D acquisitions either with or without Compressed sensing for diagnosis of iron overload and treatment monitoring, particularly if ROIs are used and are carefully placed to avoid artifacts. They can also be used interchangeably over time in the same patient. However, there are current limitations in the quality of the acquisition, particularly in-plane variation, which means it is currently not ready for use for detecting small changes in iron. In the application of iron sparing, 3D ME GRE sequences need further modification to ensure that artifacts do not impair the ability to detect small changes in R_2*. This work therefore provided the basis for further sequence optimisation which is described in the next chapter.

7. Optimisation of R_2* sequences for iron sparing detection

7.1. Introduction

Chapters 5 and 6 showed that although there are vendor sequences available for 3D R_2* mapping, there are limitations to these for the context of detecting iron sparing. For example, the standard choice of echo times is not optimised for low amounts of iron and the slice thickness can result in through-plane susceptibility which could impact the ability to detect changes in iron concentration within the liver. Additionally, Chapters 5 and 6 showed that it may be possible to use compressed sensing in low iron concentration cases but further verification of this is needed.

This chapter outlines the steps we undertook to optimise sequences at both 1.5 T and 3 T and on Philips and Siemens scanners. The optimisation was primarily performed on the Philips 3 T scanner. On the Siemens 1.5 T scanner parameters were informed by the Philips 3 T protocol but adjusted for the lower field strength. The Siemens 3 T scanner protocol parameters were matched to the Philips protocol as much as possible. Different acceleration techniques were used on both Siemens scanners compared to the Philips.

Through my collaboration with the University of Nottingham, Professor Sue Francis and Dr Eleanor Cox collected the data on the Philips 3 T scanner. All data was analysed by me, and all steps in the optimisation process discussed with me. For Siemens 1.5 T data, I collected the data alongside radiographers at Perspectum, and guided the optimisation process. I analysed the data. For Siemens 3 T data, it was initially acquired by Dr Liz Tunnicliffe at OCMR, but was subsequently collected by Dr Ferenc Mozes and myself at OCMR with me guiding the optimisation. I analysed the data.

This optimisation process was primarily performed using T_2* maps and measurements. Due to the standard practice at Perspectum (where I analysed this data) to use T_2* maps, I used T_2* throughout this work.

7.2. Materials and Methods

Three scanners were used for this work: 3 T Philips Ingenia scanner, located at the Sir Peter Mansfield Centre, University of Nottingham, 1.5 T Siemens Magnetom Aera scanner, located at Perspectum, Gemini, Oxford and a 3 T Siemens Prisma scanner, located at the Oxford Centre for Magnetic Resonance (OCMR).

Key to the optimisation process was the set of requirements given in Section 3.8.4. A baseline sequence, which could be optimised to meet the requirements, was identified on the 3 T Philips Ingenia scanner, which was the primary scanner used in this work. Given the need for a 3D ME GRE sequence, the baseline sequence chosen was mDixon Quant, the Philips product 3D T_2* mapping sequence. The baseline parameters for the mDixon Quant sequence are shown in Table 29 and Table 30. The existing vendor sequence was evaluated to determine what changes to the sequence would be needed to meet the requirements, as shown in Table 31.

	Philips 3 T, baseline parameters, mDixon Quant sequence
Field of View (mm x mm)	400 x 350
Acquisition Matrix	160 x 140
Acquired Resolution (mm x mm)	2.50 x 2.50
Reconstructed Resolution (mm x mm)	2.08 x 2.08
Acquired Slice Thickness (mm)	6.00
Reconstructed Slice Thickness (mm)	3.00
Number of slices per slab	77
Number of slabs	1
TR (ms)	5.65
Number of echoes	6

Table 29 Baseline parameters for mDixon Quant sequence, Part 1

	Philips 3 T, baseline parameters, mDixon Quant sequence
Initial TE (ms)	0.967
Final TE (ms)	4.51
Echo spacing (ms)	0.7
Bandwidth (Hz/Px)	2367
Flip Angle	3°
Parallel Imaging	SENSE 2 x 1
Compressed Sensing	N/A
Partial Fourier	N/A
Acquisition time (s)	15.5

Table 30 Baseline parameters for mDixon Quant sequence, Part 2

	Requirements for detecting Iron sparing in HCC	**Change to imaging sequence**
1)	measure iron concentration from 0 mg/g dw to 3 mg/g dw +/- 0.1 mg/g dw.	Overall length of echo train. For the study of iron sparing in HCC, the T_2* is longer than in the application of iron overload, requiring a ***longer echo train*** to fully sample the decay
2)	Be able to detect differences of 0.2 mg/g dw or more of iron concentration within the liver (for tumour identification)	SNR needs to be sufficient to detect small changes in T_2* on localised iron sparing. SNR can be maximised by ***increasing the number of echoes***, ***increasing flip angle***, and ***decreasing bandwidth***.
3)	Detection of tumours 1 cm or more.	***Increase in acquired spatial resolution in slice direction*** as 6 mm thickness may make detecting small tumours difficult
4)	Breath-hold lengths of 16 s or less, maximum of 3 breath-holds	***Increased acceleration*** using parallel imaging and compressed sensing to reduce acquisition time.
5)	Full liver coverage	Multiple slabs may be needed to give full liver coverage.
6)	Needs to be deployable on standard clinical MRI scanners without modifying their software	N/A

Table 31 Changes to the 3D ME GRE sequence to help meet the requirements for iron sparing in HCC detection

Making changes to meet a single requirement is fairly straightforward. However, when making a change, there is inevitably a trade-off which can prevent another requirement from being met. For example, increasing the spatial resolution will meet Requirement 2, but can lengthen the scan time making Requirement 4 challenging. The best trade-off must be found, and this study shows how this navigation process can be completed in practice.

To perform the optimisation, healthy volunteers (people with no known liver disease) were scanned from May 2021-April 2022. The optimisation primarily took place on the Philips 3 T scanner, but volunteers were also scanned on the other two scanners to assess the sequences developed on those scanners. All volunteers gave their informed consent to be scanned for technical development purposes. Participants scanned on the Philips 3 T scanner were scanned under Nottingham University technical development research ethics (FHMS-36-1220). The scanning of participants on the Siemens 1.5 T scanner was done under the auspices of the Clinical Affairs department at Perspectum. The participants scanned on the Siemens 3 T scanner were scanned under the OCMR Technical Development Standard Operating procedure. 11 volunteers in total are included in the evaluation given in this chapter (Philips 3 T: 5, Siemens 1.5 T: 4, Siemens 3 T: 2).

Healthy volunteers will generally have normal levels of LIC i.e. <1.8 mg/g dw[56,95], which is similar to many HCC patients[17,19,20]. Vessels were used to determine whether small structures, analogous to HCC nodules, can be detected.

Each time a volunteer was scanned on the Philips 3 T scanner various changes to the imaging sequence parameters were explored to meet the requirements, with each being assessed in turn. When a change to a parameter was made, all other parameters were kept unchanged wherever possible, so a fair comparison of the effect of the change could be made. Unless otherwise stated, the phase encode direction was anterior-posterior (y

direction) and the slice direction (which is also phase encoded) was foot-head (z direction). Each time a change was made, the improvements and limitations of the change were identified, with relation to the requirements, by viewing the T_2* maps and R^2 maps to assess the goodness-of-fit of the acquired data. A poor R^2 can indicate a reduction in quality due to poorer fitting, which may occur for a variety of reasons, for example low SNR, not sampling enough data points and artifacts. In this work R^2 maps were used primarily as an indicator of SNR. SNR was not measured directly as this is challenging, particularly when parallel imaging is used[96]. Therefore, R^2 was deemed to be a reasonable surrogate for SNR that is easy to calculate. Raw echo images were also viewed, to aid assessment of image quality.

The T_2* maps were generated offline from the individual echo images using the MAGO method [65]. T_2* maps were viewed with jet colour scale, from 0 – 50 ms. However, in this chapter, the maps are given using the Batlow colour scale, as these are accessible to those with impaired colour vision[97,98]. All slices were viewed and assessed. Generally, the images were evaluated in the axial plane, as this is the traditional plane in which radiologists view images. R^2 maps were also generated offline and viewed in greyscale, from 0.9 – 1, as well as plots of the decay curve with acquired and fitted data at a voxel level. For quantitative T_2* and R^2 evaluation, three 15 mm diameter circular ROIs were placed on a central slice avoiding vessels. The median T_2* of pixels in the ROIs calculated. An example of the maps and plots viewed is shown in Figure 32. Once an assessment of the impact of a change had been made, a decision was made as to what was the optimal solution to meet the requirements.

A similar sequence was set up on a Siemens 1.5 T scanner, with adjustments to account for the lower field strength: healthy volunteers were scanned to assess the sequence and to guide further optimisation on that scanner. As with the Philips 3 T

development, T_2* and R^2 maps were used to assess the results of changes to the sequence.

Finally, a similar sequence to the Philips 3 T sequence was set up on the Siemens 3 T scanner. Healthy volunteers were scanned to check the sequence and further changes made. As with the Philips 3 T development, T_2* and R^2 maps were used to assess the results of changes to the sequence.

7.3. Results

7.3.1. Philips Optimisation: 3 T

The results of optimisation process on the Philips 3 T scanner are given first, in the order of the requirements listed, as this was where most of the optimisation was performed.

Requirement 1: Measure iron concentration from 0 mg/g dw to 3 mg/g dw +/- 0.1 mg/g dw

Meeting Requirement 1 necessitated extending the length of the overall echo train to better fit the longer T_2* decay associated with lower levels of iron concentration in HCC patients in contrast to iron overload. The number of echoes was increased from 6 to 12 and the echo spacing extended from the shortest echo spacing of 0.7 ms to echo spacings of up to 1.6 ms. Across the range of echo spacings, T_2* values remained consistent but the longer echo spacing gave a better fit as shown by the higher R^2. An example of increasing the echo spacing is shown in Figure 32.

Meeting Requirement 1 is fundamental; but inevitably increases the acquisition time, as longer echo trains typically require a longer TR. This makes meeting Requirement 4 challenging. For example, in Figure 32, the breath-hold length for the shorter echo train was 7.79 s for 45 axial slices, whereas for the longer echo train it was 16.5 s with 43 axial

slices. Additionally, when the echo spacing is longer, the fit will worsen for patients with higher iron concentration, i.e., shorter T_2*.

Outcome: This first phase of optimisation resulted in a 12 echo train with an echo spacing of 1.5 ms and thus an echo train of 1-17.5 ms.

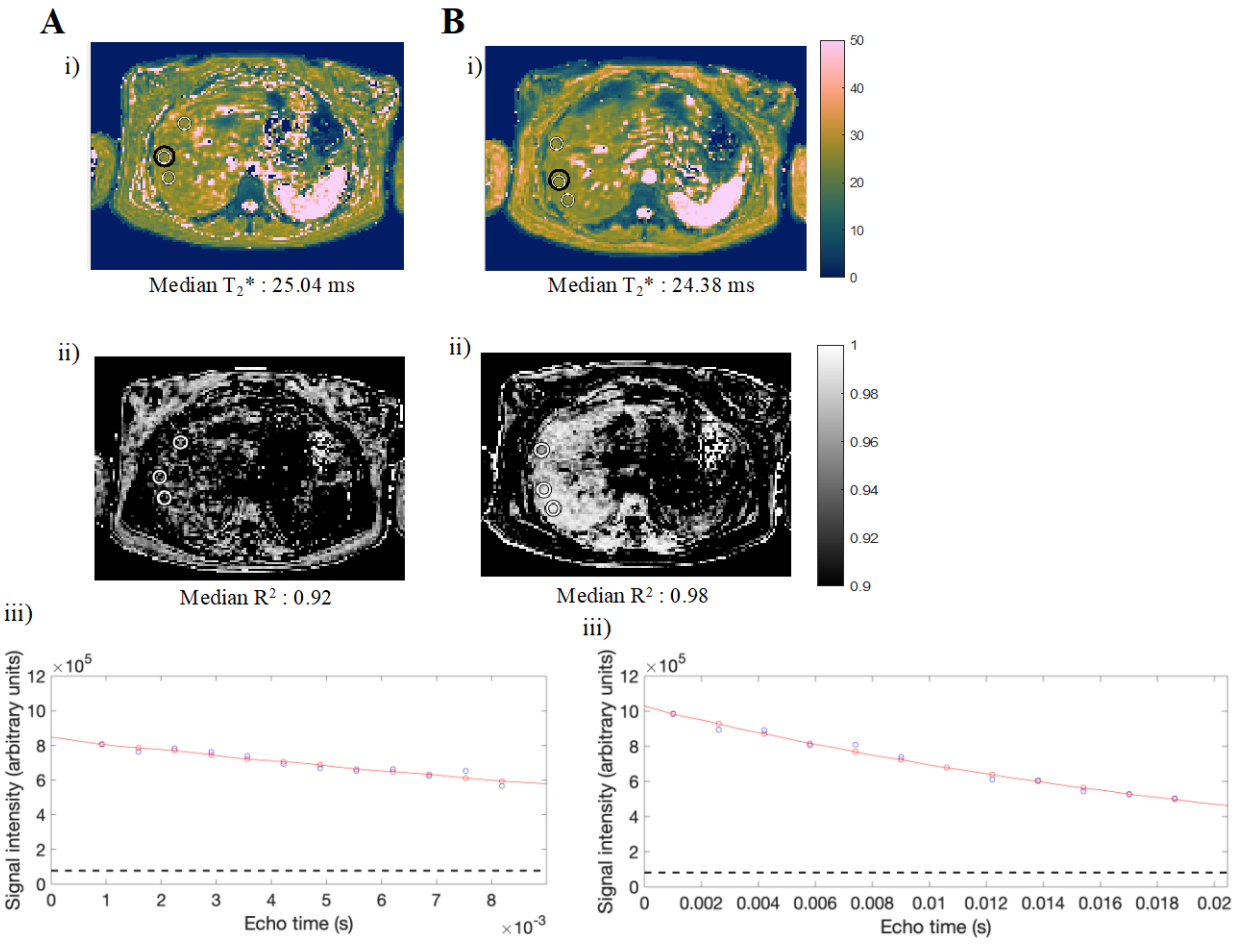


Figure 32 i) T_2* map and ii) R^2 map for (A) short (0.922 - 8.191 ms) and (B) long echo train (1 - 18.601 ms). Three ROIs were assessed indicated by the white circles on the maps with the median of these ROIs given below each map and iii) the fit of the ROI circled in black shown. The acquired data is given by the blue circles, the fitted data by the red circles, and the fit of the T_2* decay curve by the red line. The dashed line is the noise cut off. The fit can oscillate due to the presence of fat. The R^2 map, R^2 median and plot of the fit of a voxel all show improvement for longer echo spacing.

Requirement 2: Detect differences of 0.2 mg/g dw or more of iron concentration within the liver (for tumour identification)

To meet Requirement 2, changes to maximise SNR were explored and evaluated visually on the T_2* maps and from the R^2 maps. Increasing the number of echoes from 6 to 12 to meet Requirement 1 was found to have improved the SNR as reflected in the less noisy T_2* map (Figure 33). Additionally, R^2 is higher for 12 echoes than 6, which is a strong result, given that, all things being equal, we expect a drop in R^2 when fitting with more data points.

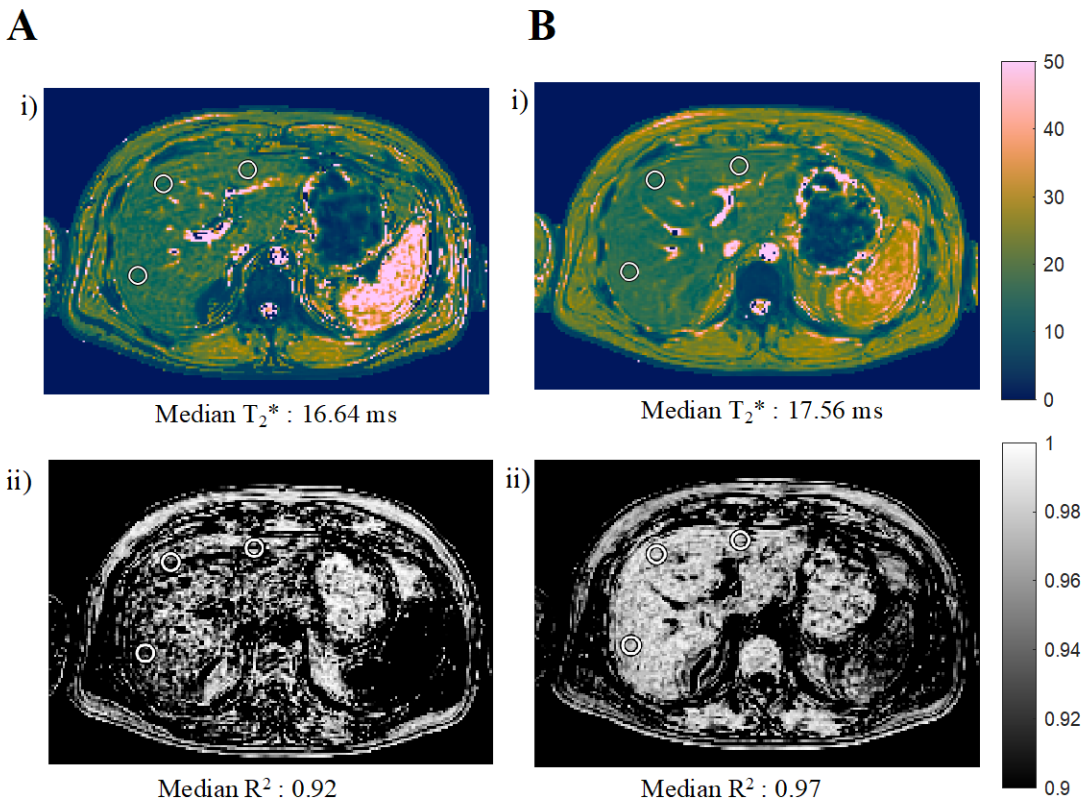


Figure 33 i) T_2* map and ii) R^2 map for (A) 6 echo (0.967-4.51ms) and (B) 12 echo train (0.997-9.086ms). Echo spacing was kept similar for both: around 0.7 ms. Three ROIs were assessed indicated by the white circles on the maps with the median of these ROIs given below each map. R^2 is higher for a longer echo train.

A further way to improve the SNR is to decrease the bandwidth, which is typically high on ME GRE sequences. This did not have any noticeably negative effects on meeting the requirements, however a reduced bandwidth can result in chemical shift artifact,

particularly if there is more fat in the tumour than the parenchyma (or vice versa). A third way to improve the SNR is to increase the flip angle. This resulted in improved SNR (assessed visually and using R^2 to assess goodness-of-fit), and no immediately obvious negative effects in meeting the requirements. However, it is known that flip angle can cause T_1 bias. A final way to boost SNR is to lower the spatial resolution, but this would have had a direct negative implication in satisfying Requirement 3.

Outcome: 12 echoes were used. A bandwidth of 1480 Hz/Pixel was chosen. The flip angle was kept low at 3°.

Requirement 3: Detect tumours of size 1 cm or more

To meet Requirement 3 the spatial resolution was changed. This is often low on standard ME GRE sequences in the through-plane (z) direction which could result in not detecting a tumour. Therefore, first, the voxel resolution in the z direction was changed, with the aim of making a more isotropic sequence. As shown in Figure 34, as the voxel size was decreased, the vessel delineation improved. However, SNR decreased, evident in the median R^2 reducing from 0.97 to 0.94 when acquired resolution was modified from 2.5 x 2.5 x 6 mm to 2 x 2 x 3 mm. This could impact meeting Requirement 2.

When changing the voxel dimension in the z direction the number of slices was kept the same, so the acquisition time was not impacted, but the coverage was reduced, due to the smaller voxel sizes. This therefore impacts Requirement 4, as either longer or more breath-holds are required to obtain full liver coverage. Additionally, when the resolution was increased in the x and y directions the field of view in the x and y directions was kept the same, so this led to an increase of breath-hold length, with implications for Requirement 4.

Outcome: Given the impact of increasing resolution on SNR and acquisition time, it was decided to acquire at 3.00 x 3.00 x 3.00 mm, reconstructed to 2.75 x 2.75 x 1.50 mm.

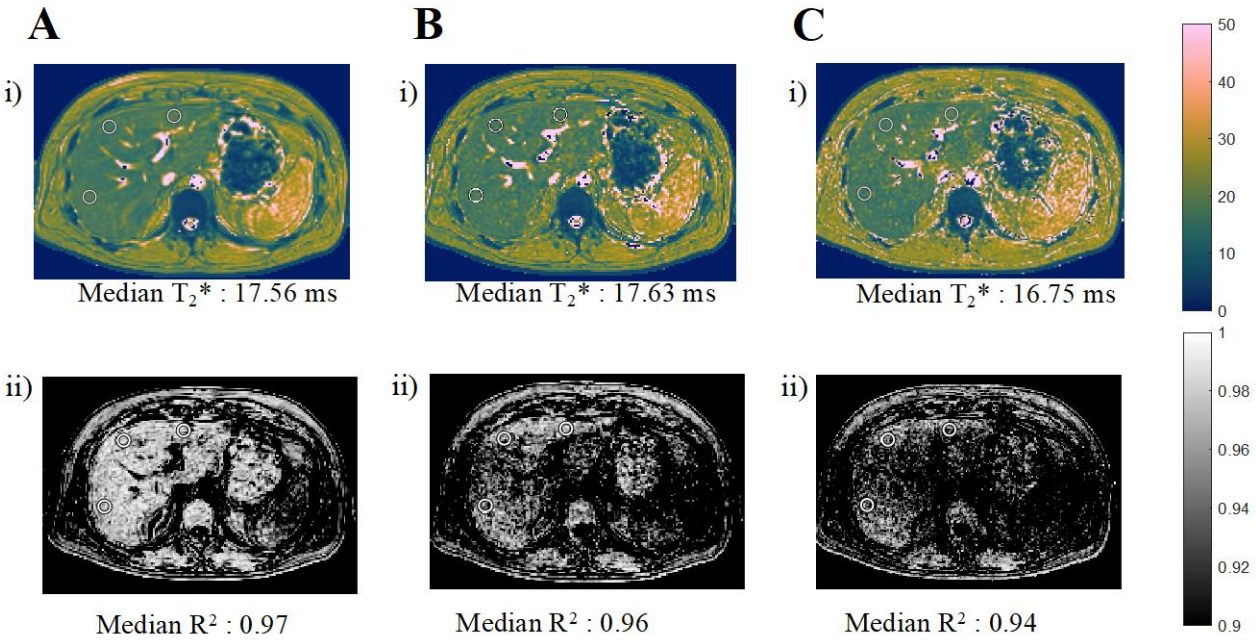


Figure 34 i) T_2* map and ii) R^2 map for a 12 echo sequence with an echo train length of 9-10 ms for different resolutions: (A) 2.5 x 2.5 x 6 mm reconstructed to 2.08 x 2.08 x 6 mm, (B) 2.5 x 2.5 x 3 mm reconstructed to 2.0833 x 2.0833 x 3 mm and (C) 2 x 2 x 3 mm reconstructed to 1.7857 x 1.7857 x 3 mm. Three ROIs were assessed indicated by the white circles on the maps with the median of these ROIs given below each map. R^2 is higher for a larger voxel volume.

Requirement 4: Scan breath-hold length of 16 s or less, maximum of 3 breath-holds

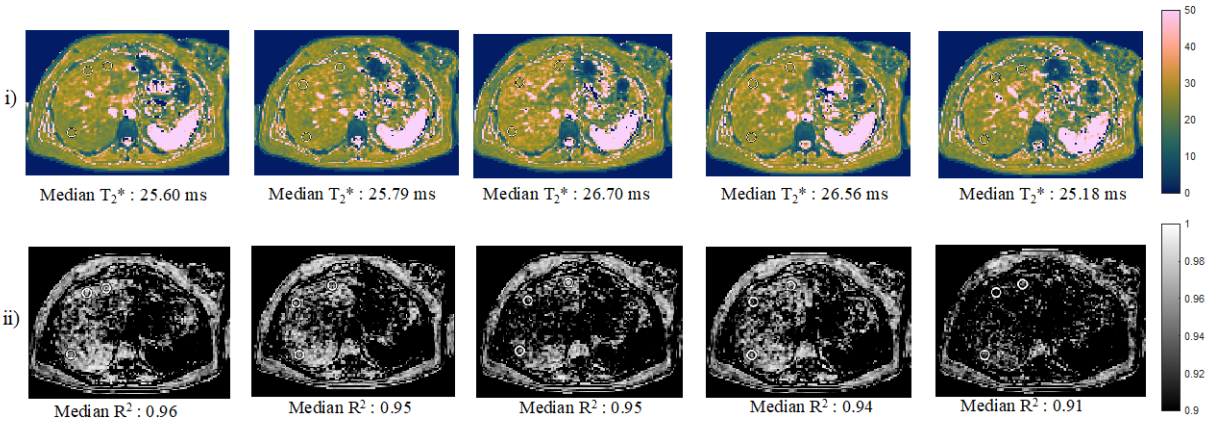
Requirement 4 was the most challenging requirement to meet, given the other changes lengthened the acquisition. To achieve this, the use of both parallel imaging and compressed sensing was explored.

Parallel Imaging:

On the Philips systems parallel imaging is performed using SENSE, for 3D imaging the user specifies SENSE factors (i.e., acceleration factors) in the phase and slice direction. Figure 35 shows the impact of increasing SENSE factor on increasing the number of slices and/or decreasing the acquisition time to meet Requirement 4, with a resultant T_2* map of the central slice of the liver shown. This resulted in a decrease in SNR, evident from the lower R^2 maps. Although not shown in the figure, this was

particularly found to be the case in the uppermost and lowermost slices in the slab, hence making Requirement 2 more difficult to meet.

Outcome: SENSE 2 (phase direction) and 1.7 (slice direction) provided best balance between acceleration and SNR, resulting in a breath-hold length of 16.4 s and 44 slices per stack.



SENSE phase direction	2	2.5	3	2	2
SENSE slice direction	1	1	1	1.5	2
Number of slices	63	75	89	83	89
Scan duration (s)	17.5	16.7	16.4	16.6	13.9

Figure 35 i) T_2* map and ii) R^2 map for a 12 echo sequence with an echo train length of 8.191 ms for different SENSE factors in phase and slice directions. Three ROIs were assessed indicated by the white circles on the maps with the median of these ROIs given below each map. R^2 is higher for lower SENSE factors.

Compressed Sensing

Another way to reduce the imaging time is compressed sensing. The scanner in this study had compressed sensing available. We increased the compressed SENSE factor from 1.87 to 6.5. The quality remained good, with R^2 of 0.98 at compressed SENSE 6.5, and T_2* values remained consistent. See Figure 36 for an example of increasing Compressed SENSE factor. It resulted in much shorter breath-hold lengths for more slices.

Outcome: If compressed SENSE is available use compressed SENSE 6.5 with a breath-hold length of 12.9 for 64 slices.

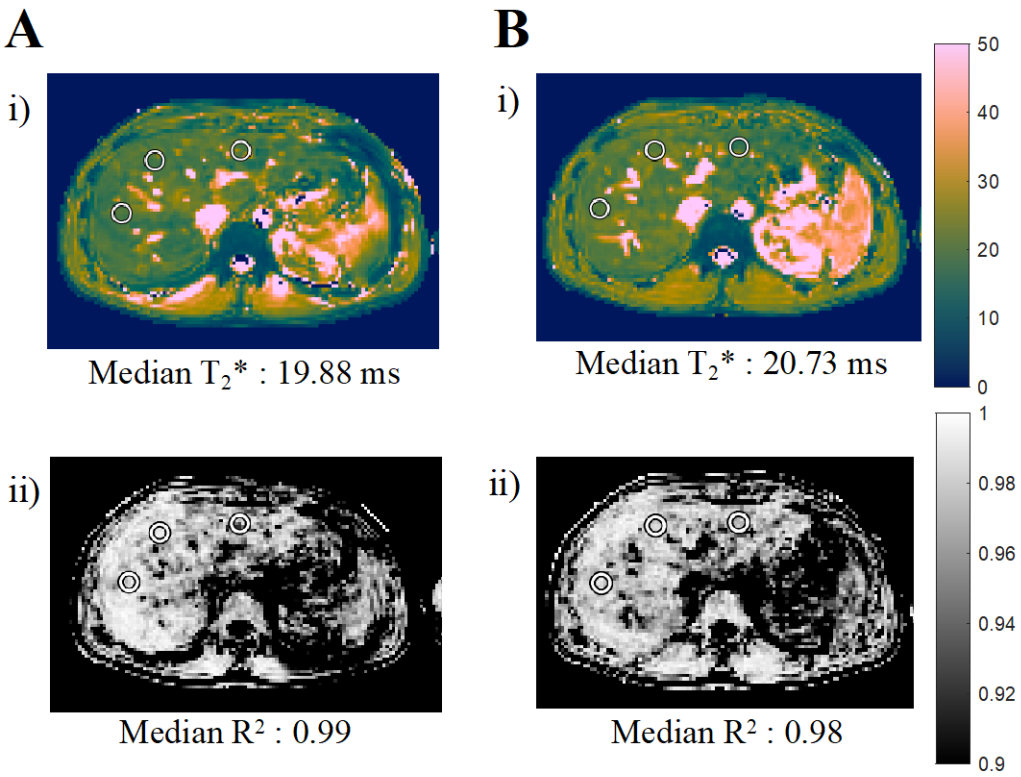


Figure 36 i) T_2* map and ii) R^2 map for a 12 echo sequence with an echo train length of 17.5 ms for compressed SENSE factors of (A) 4.5 and (B) 6.5. Three ROIs were assessed indicated by the white circles on the maps with the median of these ROIs given below each map. R^2 is high for both compressed SENSE factors.

Requirement 5: Full liver coverage

The next requirement to meet was Requirement 5. The baseline sequences were 3D, but once changes were made to fulfil Requirements 1-3, they did not give full coverage of the liver, and so the sequence had to be collected in at least two slabs for whole liver coverage. An overlap was required as quality decreases at the slab edges, as manifested by the noisy appearance of the T_2* maps and poor R^2. This was particularly the case for the acquisition using SENSE, as opposed to compressed SENSE. For compressed SENSE the quality improved sufficiently by around 5 mm below the top for quantification, whereas for SENSE the quality remained impaired until around 10 mm below the top. Examples of T_2* maps and R^2 maps at the edges of a slab are shown in Figure 37 and

Figure 38 and Given the impaired quality of slices at the periphery of the slaps, an overlap of 20 mm was decided on for SENSE and an overlap of 10 mm for the Compressed SENSE acquisition.

Outcome: SENSE acquisition use 2-3 slabs with overlap of 20 mm, with compressed SENSE use 2-3 slabs with overlap of 10 mm.

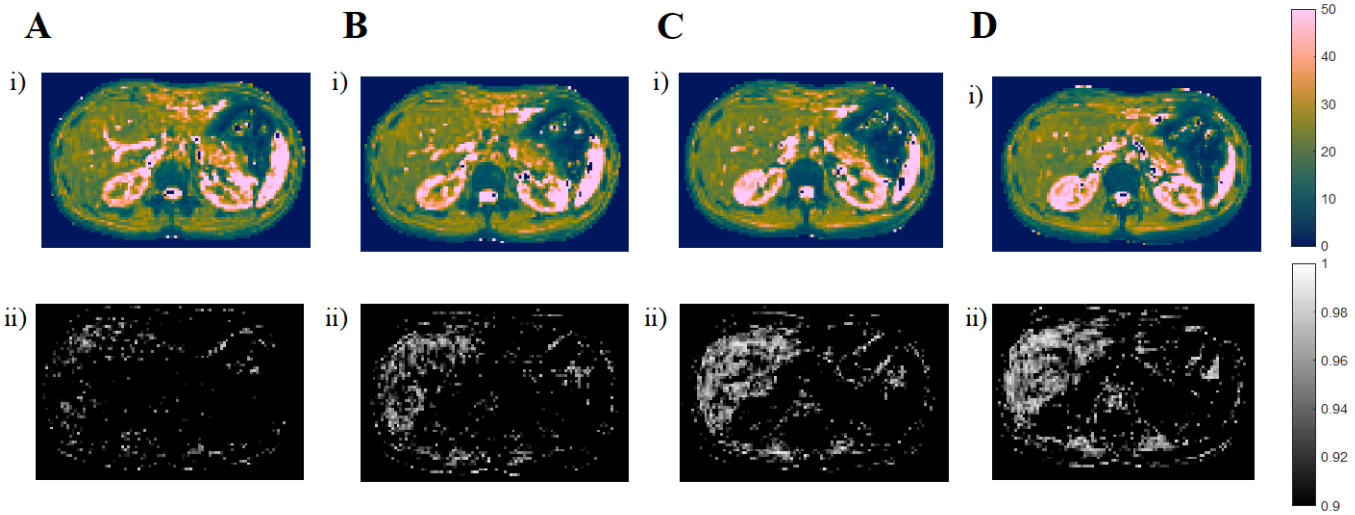


Figure 37 i) T_2* map and ii) R^2 map with the final parameters for Philips 3 T acquisition with Compressed SENSE, showing slices in the lower slab located at (A) top of lower slab (B) 3 mm below top slice (C) 6 mm below top slice (D) 9 mm below top slice. SNR is worse on the uppermost slices but by 6 mm below the top slice the SNR is better, as seen from the T_2* map and R^2 map.

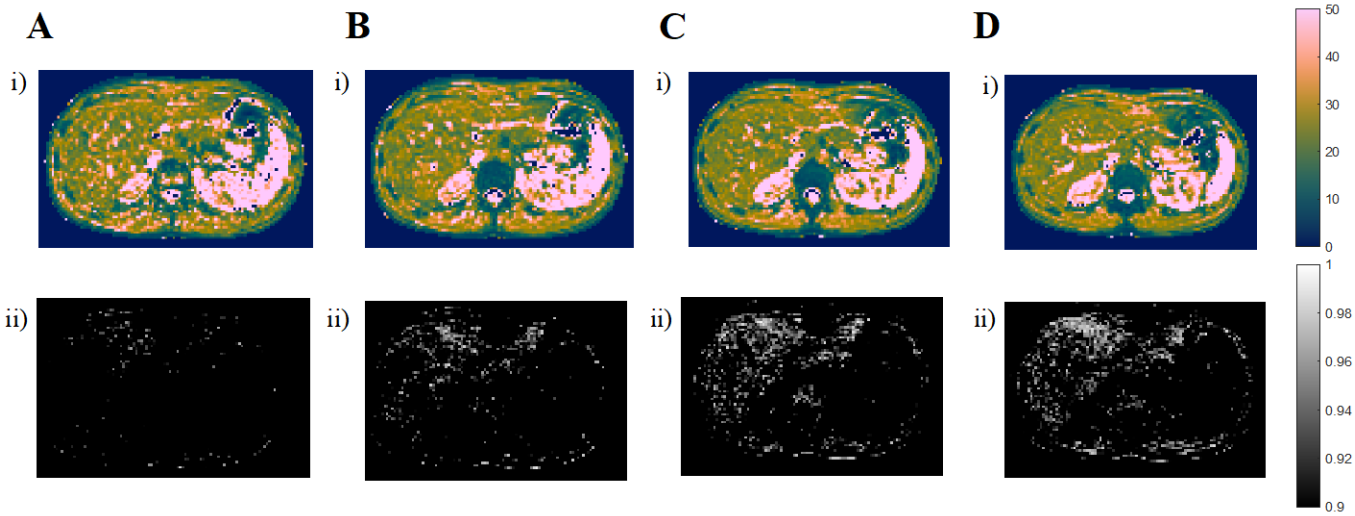


Figure 38 i) T_2* map and ii) R^2 map with the final parameters for Philips 3 T acquisition with SENSE, showing slices in the lower slab located at (A) top of lower slab (B) 3 mm below top slice (C) 6 mm below top slice (D) 9 mm below top slice. SNR is worse on the periphery of the slice and does not improve enough for quantification until around 9 mm below the top of the slab, as shown by the T_2* maps and R^2 maps.

Requirement 6: Needs to be deployable on standard clinical MRI scanners without modifying their software

Requirement 6 was met, as no modification of the standard clinical MRI scanners used was made, and no pulse programming was required to alter the sequence parameters.

7.3.2. Siemens Optimisation: 1.5 T

Following the changes made to the 3D ME GRE sequence on the Philips 3 T scanner, similar sequences were set up on the Siemens 1.5 T and 3 T scanners. At 1.5 T changes were needed to accommodate for the slower T_2* decay and for the decrease in signal at 1.5 T.

Firstly, given that T_2* decay is slower at 1.5 T compared to 3 T, the decision was made to have a larger echo spacing of 2.49 ms, with echo train 1.59 – 28.98 ms to meet Requirement 1. Given the lower signal at 1.5 T, to boost SNR to help with meeting Requirement 2, the bandwidth was dropped to 550 Hz/Px. The flip angle was increased to 12°, as this also increases the signal. It is known that flip angle can cause T_1 bias[99]. However, the impact of T_1 bias is mainly on the fat fraction (which can also be calculated from ME GRE sequences) and not on T_2*[100] and so, given that the primary purpose of this sequence was for T_2* quantification, this was not a major concern. If the sequence was used for fat quantification then potentially a T_1 correction could be applied to account for the higher flip angle[99]. A final way to boost SNR is to increase the voxel size. This results in a trade off with Requirement 3 (to detect tumours of 1 cm or more), but a compromise was made with acquired resolution 3.96 x 3.96 x 4.00 mm, reconstructed resolution 2.00 x 2.00 x 2.00 mm. Having the larger voxel size also helped with meeting Requirement 4 as it reduces the scan time, especially important given that the longer TEs for the 1.5 T sequence extended the acquisition time.

As well as changes specifically related to being at 1.5 T, other changes were needed compared to the Philips scanner as the parallel imaging techniques are different on Siemens scanners and Compressed Sensing was not available on either Siemens scanner.

Siemens has two implementations of parallel imaging: GRAPPA and CAIPIRINHA. CAIPIRINHA is best suited for 3D liver imaging (see Section 3.5.2). Like Philips, the amount of acceleration applied is changed by increasing the acceleration factor, the iPAT factor. Unlike Philips, only integer factors can be used. Increasing the iPAT factor can cause artifacts. Firstly iPAT 2 in both phase and slice directions (i.e., overall factor 4) on the 1.5 T scanner was assessed empirically, which had breath-hold length of 16 s. However, this resulted in artifacts in some volunteers, see Figure 39. These were also evident on the raw echo images. The impact on the T_2* map could impair the ability to detect iron differences (Requirement 2). When iPAT 3 in the slice direction only was used the artifact was not present, and so therefore it was decided that this would be the best acceleration factor to use. However, the breath-hold length was 20 s, so did not meet Requirement 4.

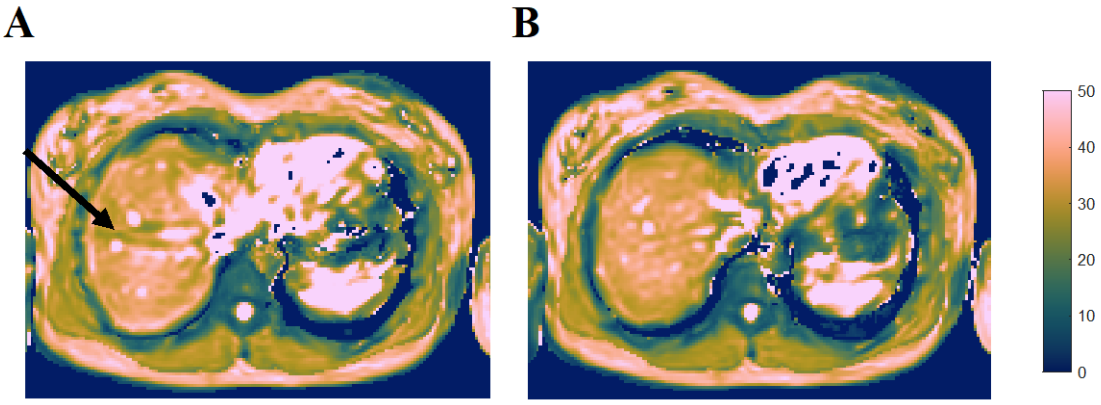


Figure 39 Volunteer scanned on Siemens 1.5 T, T_2* maps: (A) iPAT 2 phase direction and iPAT 2 slice direction, parallel imaging artifact indicated by arrow; (B) iPAT 1 phase direction, iPAT 3 slice direction

Given that on the Siemens 1.5 T scanner the breath-hold length was still too long, using Partial Fourier was explored. Partial Fourier phase direction 7/8 and 6/8 and slice direction 7/8, 6/8 and 5/8 were evaluated, and the various combinations of these. The

respective breath-hold lengths of each of the changes to the Siemens 1.5 T sequence (with iPAT 3) are given in Table 32. By combining Partial Fourier in the phase and slice directions reduced the acquisition time further.

Partial Fourier	Breath-hold length (s)
None	20
Phase 7/8	18
Phase 6/8	16
Slice 7/8	18
Slice 6/8	16
Slice 5/8	15

Table 32 Impact of different amounts of Partial Fourier on Breath-hold length (1.5 T)

However, although Partial Fourier substantially reduced acquisition time, helping to satisfy Requirement 4, vessels became less clear and of lower T_2* with increased Partial Fourier (Figure 40). This is concerning for meeting Requirement 3, if small structures cannot be seen. Additionally, T_2* was shortened near the lungs (Figure 41), which could impact meeting Requirement 1 and Requirement 2. It was decided to use Partial Fourier 7/8 in phase and slice direction to provide a compromise between accelerating the acquisition and limiting the negative effects of using Partial Fourier. The length of breath-hold was 16 s, which met Requirement 4, whilst not impacting achievement of the other requirements.

Like the Philips 3 T sequence, multiple slabs are required to obtain full liver coverage. After evaluating the quality at the edges of the slab it was decided to have an overlap of 10 mm for the Siemens 1.5 T sequence.

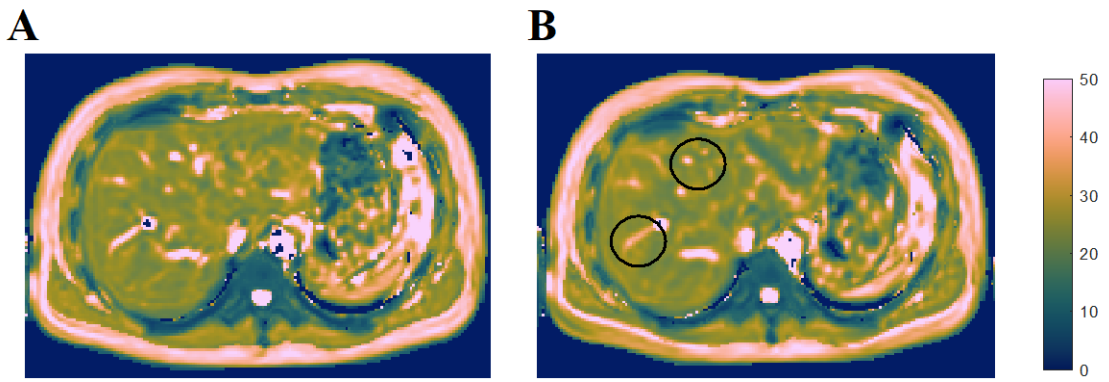


Figure 40 Volunteer scanned on Siemens 1.5 T, T_2* maps: (A) No Partial Fourier, (B) Partial Fourier 6/8 Phase direction: examples of vessels which are less apparent/of shorter T_2* are circled

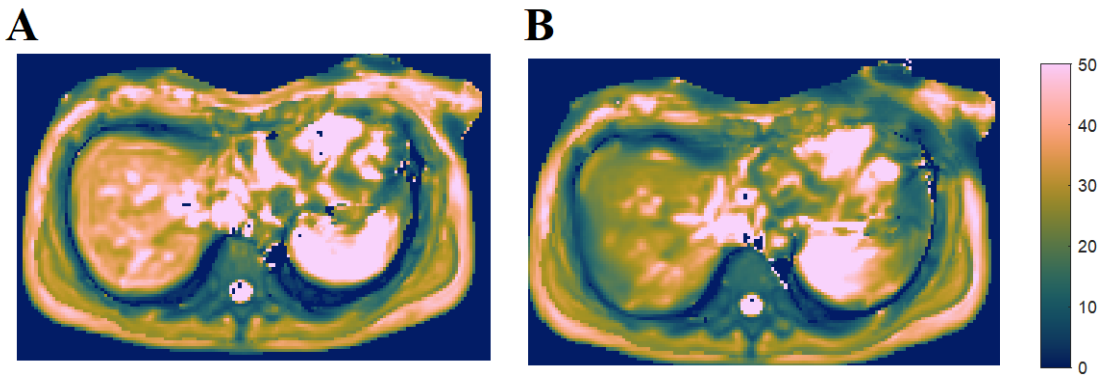


Figure 41 Volunteer scanned on Siemens 1.5 T, T_2* maps: (A) No Partial Fourier, (B) Partial Fourier 5/8 Slice direction. T_2* is shorter with Partial Fourier 5/8.

7.3.3. Siemens Optimisation: 3 T

For the Siemens 3 T protocol, the aim was to match the Philips 3 T protocol as closely as possible. Some changes were needed as the scanner did not always permit identical parameters. For example, the initial TE had to be 1.45 ms compared to 1 ms and the bandwidth had to be 1560 Hz/Px compared to 1482 Hz/Px.

The Siemens 3 T scanner had iPAT and CAIPIRINHA. Given the findings on Siemens 1.5 T (and prior experience at 3 T), I decided to use the same iPAT factor as at 1.5 T: 1 (phase encode direction) x 3 (slice direction). Partial Fourier was required to reduce the breath-hold length to meet Requirement 4. However, for Partial Fourier 7/8 in the slice direction, the slices near the dome of the liver had T_2* shortening compared to no Partial Fourier, as shown in Figure 42. The impact of susceptibility is larger at 3 T compared to 1.5 T; hence the T_2* shortening effect is worse. Therefore, instead of using Phase and

Slice Partial Fourier 7/8 as for 1.5 T, it was decided to try using Phase Partial Fourier 6/8. The impact of using Partial Fourier 6/8 on delineation of the vessels was less than at 1.5 T, see Figure 43, and so it was decided to use Partial Fourier 6/8. This resulted in a breath-hold length of 15 s, thus meeting Requirement 4. For coverage of the whole liver (Requirement 5) at least 3 slabs. An overlap of 20 mm was used, as this matches the Philips 3 T overlap for the sequence without compressed sensing. No modification of the scanner software was required, thus meeting Requirement 6.

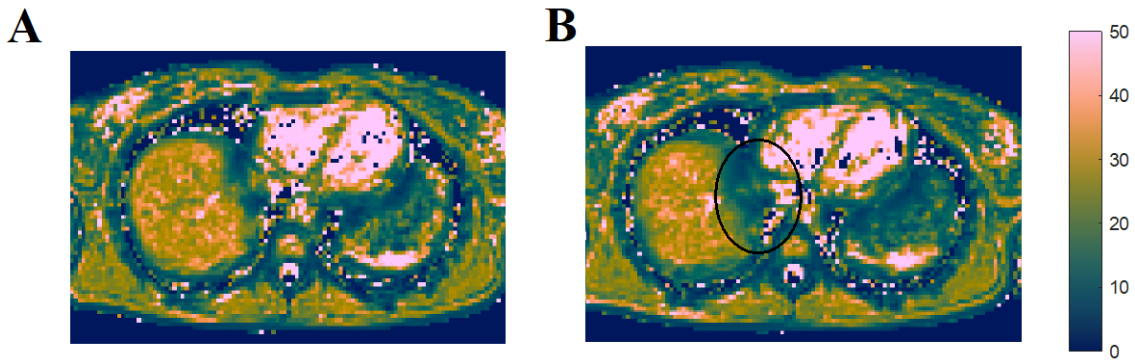


Figure 42 Volunteer scanned on Siemens 3 T (A) No Partial Fourier, (B) Slice Partial Fourier 7/8. When Slice Partial Fourier is applied there is T_2* shortening close to the heart (black circle), which is less with no Partial Fourier.

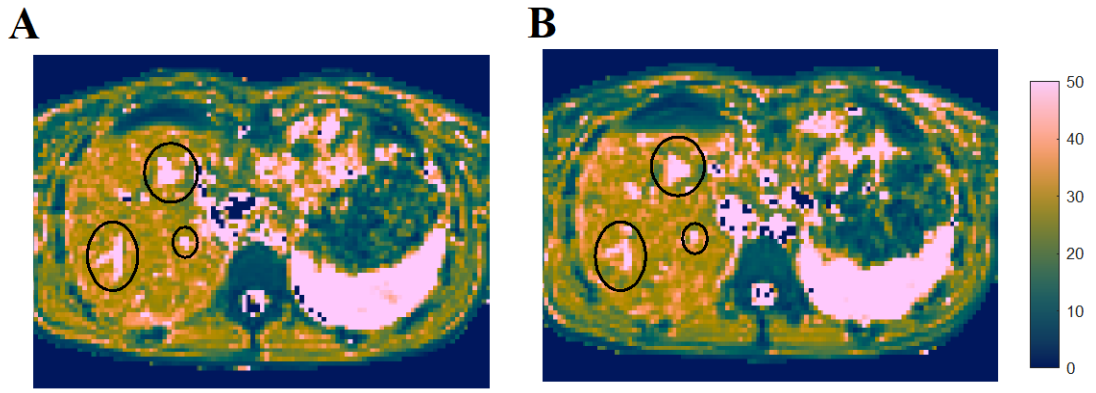


Figure 43 Volunteer scanned on Siemens 3 T (A) No Partial Fourier, (B) Phase Partial Fourier 7/8. When using Phase Partial Fourier, the same vessels can still be seen as when no Partial Fourier is used (black circles around vessels)

7.3.4. Final Protocol

A summary of the changes against the requirements is given in Table 33 and final parameters are given in Table 34. T_2* maps using the final sequences are shown in: Figure 44 for Philips 3 T with SENSE, Figure 45 for Philips 3 T with compressed SENSE, Figure 46 for Siemens 1.5 T and Figure 47 for Siemens 3 T.

Requirements for detecting Iron sparing in HCC	Change to imaging sequence	Philips 3 T	Siemens 1.5 T	Siemens 3 T
1) measure iron concentration from 0 mg/g dw to 3 mg/g dw +/- 0.1 mg/g dw.	Overall length of echo train. For the study of iron sparing in HCC, the T_2* decay is longer than in the application of iron overload, requiring a ***longer echo train***.	12 echoes, with echo train 1-17.5 ms.	12 echoes, like 3 T, but at T_2* decay is slower than 3 T, so decision was made for an echo spacing of 2.49 ms, with echo train 1.59 – 28.98 ms	12 echoes, with echo train 1.45-17.5 ms.
2) Be able to detect differences of 0.2 mg/g dw or more of iron concentration within the liver (for tumour identification)	SNR needs to be sufficient to detect small changes in T_2* on localised iron sparing. SNR can be maximised by ***increasing the number of echoes***, ***increasing flip angle***, and ***decreasing bandwidth***.	12 echoes Bandwidth 1482 Hz/Px	12 echoes Bandwidth 550 Hz Flip angle 12° (to boost signal, as signal is less at 1.5 T)	12 echoes Bandwidth 1560 Hz/Px
3) Detection of tumours 1 cm or more.	***Spatial resolution must be increased*** such that that 1 cm tumours can be detected.	Acquired resolution 3.00 x 3.00 x 3.00 mm, reconstructed resolution 2.75 x 2.75 x 1.50 mm	Acquired resolution 3.96 x 3.96 x 4.00 mm, reconstructed resolution 2.00 x 2.00 x 2.00 mm (larger voxel sizes compared to 3 T to boost SNR given lower signal intensity and speed up acquisition given longer TEs)	Acquired resolution 2.75 x 2.84 x 3.00 mm reconstructed resolution 2.75 x 2.75 x 1.50 mm
4) Breath-hold lengths of 16 s or less, maximum of 3 breath-holds	***Increased acceleration*** using parallel imaging and compressed sensing to reduce acquisition time.	Use SENSE parallel imaging or Compressed SENSE	Use iPAT and Partial Fourier (Compressed sensing not available on scanner)	Use iPAT and Partial Fourier (Compressed sensing not available on scanner)
5) Full liver coverage	Multiple slabs may be needed to give full liver coverage.	Use up to 3 slabs	Use up to 3 slabs	Use up to 3 slabs
6) Needs to be deployable on standard clinical MRI scanners without modifying their software	N/A	Met	Met	Met

Table 33 Summary of sequence modifications to optimise ME GRE sequences for detecting iron sparing

Parameter	Philips 3 T, SENSE version	Philips 3 T, Compressed SENSE version	Siemens 1.5 T	Siemens 3 T
Field of View (mm)	440 x 374	440 x 374	380 x 333	440 x 374
Matrix size	148 x 124	148 x 125	96 x 84	160 x 109
Acquired Resolution (mm)	2.97 x 3.02	2.97 x 2.99	3.96 x 3.96	2.75 x 3.44
Reconstructed Resolution (mm)	2.75 x 2.75	2.75 x 2.75	2.00 x 2.00	2.75 x 2.75
Acquired Slice thickness (mm)	3.00	3.00	4.00	3.00
Reconstructed Slice thickness (mm)	1.5	1.5	2.00	1.5
Number of slices per slab	44	64	30	40
Slab Overlap (mm)	20	10	10	20
TR (ms)	18.77	18.77	31.80	19.00
Number of echoes	12	12	12	12
Initial TE (ms)	1.00	1.00	1.59	1.45
Final TE (ms)	17.5	17.5	28.98	17.51
Echo spacing (ms)	1.5	1.5	2.49	1.46
Bandwidth (Hz/Px)	1482	1482	550	1560
Flip Angle	3°	3°	12°	3°
Parallel Imaging	SENSE Factor 2 (phase encode direction) x 1.7 (slice direction)	N/A	iPAT (CAIPIRINHA): 1 (phase encode direction) x 3 (slice direction)	iPAT (CAIPIRINHA): 1 (phase encode direction) x 3 (slice direction)
Compressed Sensing	N/A	Compressed SENSE 6.5	N/A	N/A
Partial Fourier	N/A	N/A	Phase encoding direction 7/8, Slice direction 7/8	Phase encoding direction 6/8
Acquisition time (s)	16.4	12.9	16	15

Table 34 Final parameters of the optimised sequences

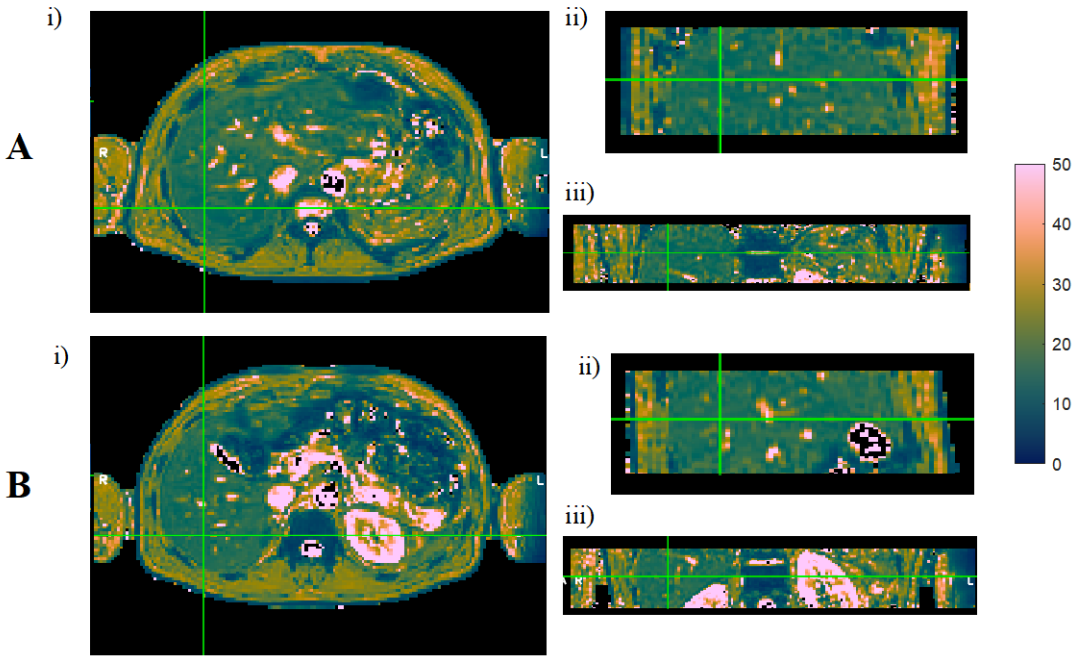


Figure 44 Philips 3 T T_2* maps for the (A) upper and (B) lower slabs in the i) axial, ii) sagittal and iii) coronal views for a SENSE factor of 2 x 1.7.

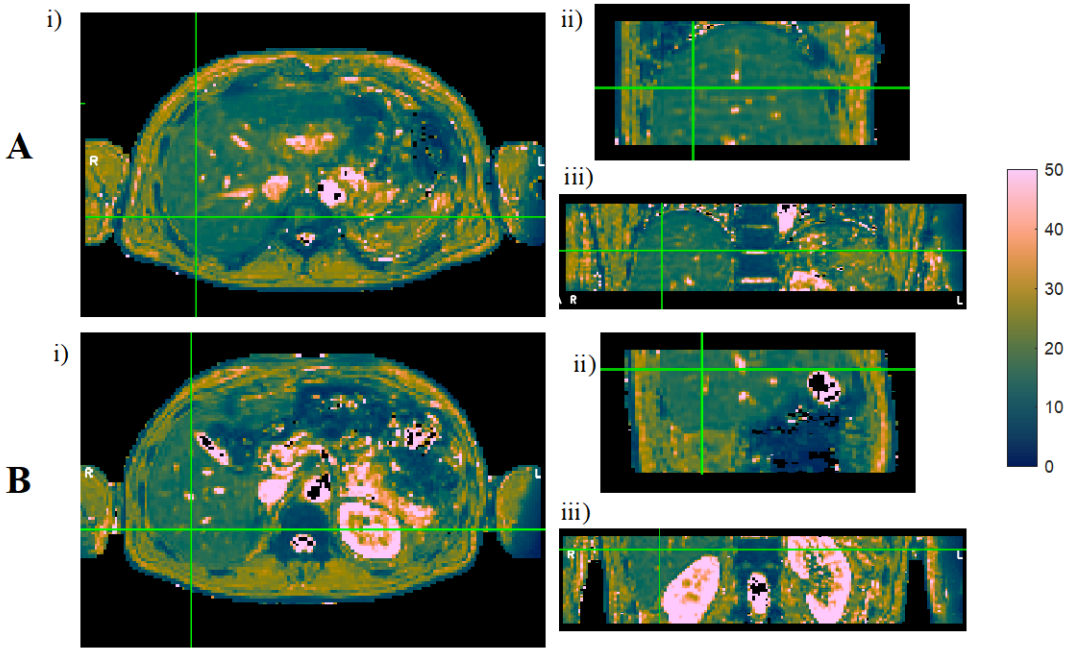


Figure 45 Philips 3 T T_2* maps for the (A) upper and (B) lower slabs in the i) axial, ii) sagittal and iii) coronal views for a compressed SENSE factor of 6.5.

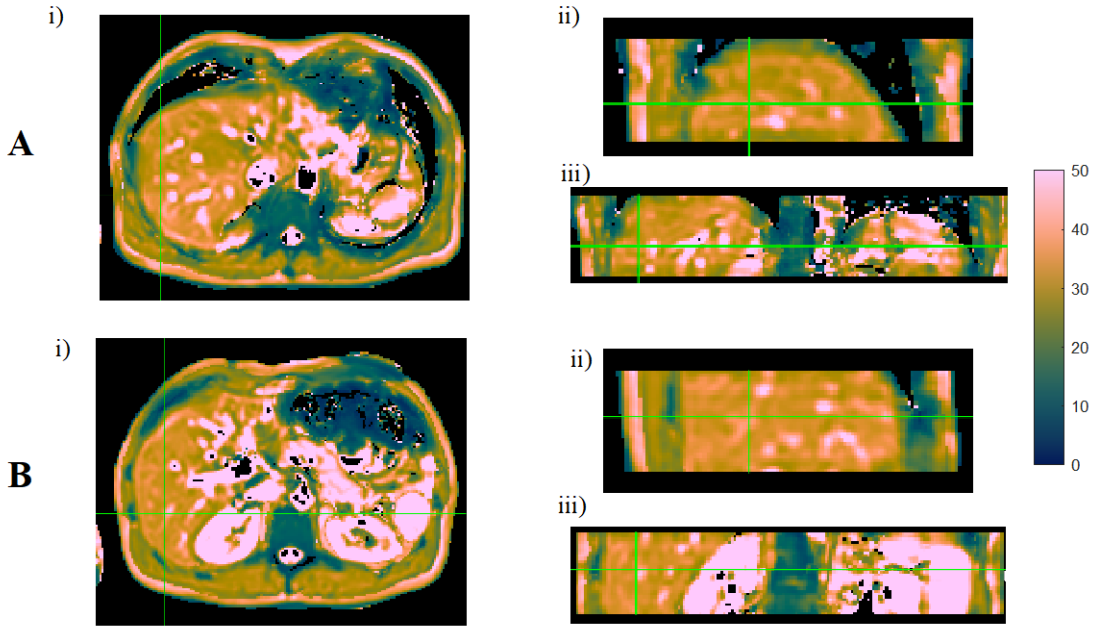


Figure 46 Siemens 1.5 T T_2* maps for the (A) upper and (B) lower slabs in the i) axial, ii) sagittal and iii) coronal views.

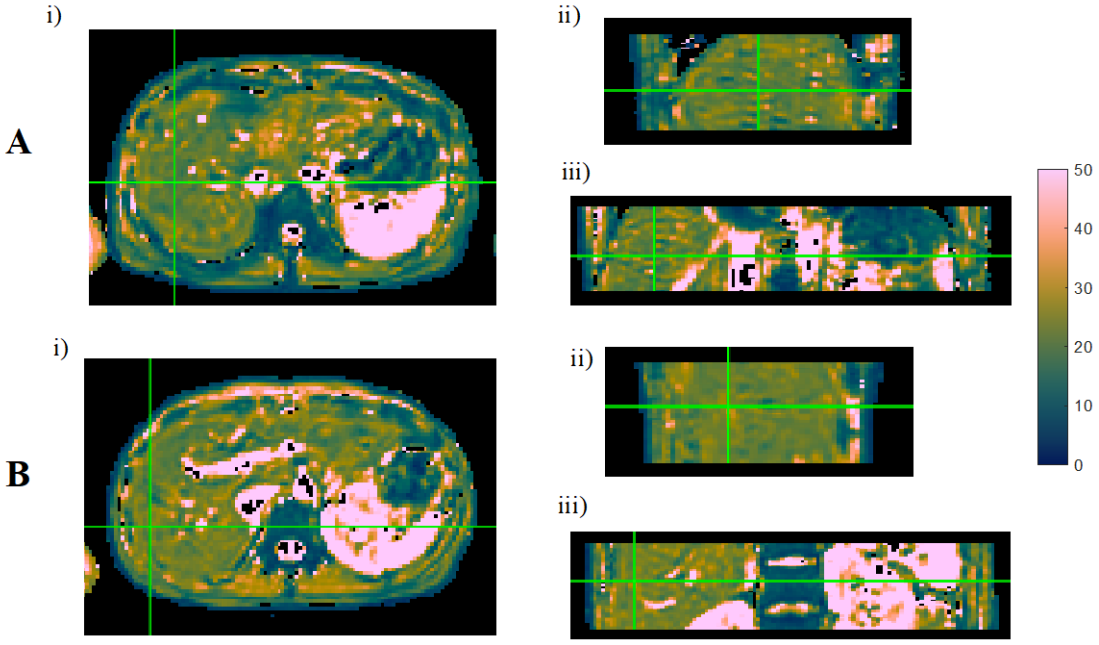


Figure 47 Siemens 3 T T_2* maps for the (A) upper and (B) lower slabs in the i) axial, ii) sagittal and iii) coronal views.

7.4. Discussion

This chapter reports on a detailed design process for optimising a 3D ME GRE sequence for use in the study of T_2* maps for HCC detection. The process necessitated careful navigation between a set of conflicting requirements to arrive at the final sequence. A key initial step was to formalise requirements. Without these, it would be difficult to assess whether adjustments were making the sequence more suitable for HCC detection. A suitable sequence, and then potential changes to this sequence to meet the requirements, were identified. It was important to identify the changes and establish in what order the requirements should be addressed, so a structured and methodical approach could be taken. Then the changes were made, and the results evaluated against the requirements. Similar sequences were developed on scanners made by a different manufacturer, including one at a different field strength, and the sequences evaluated.

One important change made was to increase the echo spacing, so the length of the echo train was increased. This resulted in an improvement of the fit in these healthy volunteers, which we would also expect in patients at-risk of HCC who are likely to have similar levels of iron to the normal iron concentration range[17,19,20,56]. However, it should be noted that some patients at-risk of HCC will have higher iron[101], and in such cases the longer echo spacing could have negative impacts: if the decay is too fast the fit may not be estimated correctly due to limited number of echoes whilst the decay is taking place. Therefore, it may be appropriate to have an additional sequence with shorter echo spacing so that this screening tool can be used in all patients. Lengthening the echo train can result in increased scan time (as it can result in increasing TR). This was particularly a challenge at 1.5 T where the echo train was lengthened more due to the slower signal decay, and so the preference of scanner for screening of this type is 3 T.

The methods used to increase SNR worked well, as was manifested by the improved R^2 maps. However, awareness of the trade-offs, though not manifested in resulting images, should be considered; firstly, decreasing bandwidth can cause chemical shift artifacts[102], and secondly, increasing the flip angle can result in T_1 bias [103].

The most challenging part of the design process was shortening the acquisition time. This also highlighted the differences between scanner manufacturers. Implementing SENSE parallel imaging led to the confounding factor of a decrease in SNR, particularly noticeable at the edges of the slab imaged. Particular attention needs to be paid here as this means that a substantial overlap is needed between slabs to ensure good quality T_2* maps are available throughout the liver. With Siemens' CAIPIRINHA, the main challenge was the artifacts that it caused. These were not always present, and sometimes were only obvious from the raw images. However, this means that it is not appropriate to use high iPAT factors, as even if artifacts are not present in the maps for every patient, they could limit the detection of HCC where they are present. The fact that they may not be evident in the maps but are in the raw images is concerning: the reviewer of the images may not realise that the maps are impaired, but even if not obvious in the maps, it could still make detection of a tumour in the region impacted challenging. Another limitation of the Siemens scanner, relative to the Philips scanner, was that iPAT only allows integer values.

Partial Fourier also has limitations in this application. For example, when applied in the slice direction it can cause T_2* shortening near the lungs. This is because the phase correction used in the Partial Fourier process is based on the slowly varying phase shifts, which does not apply to the large phase change between the liver and lungs. Additionally, it can reduce the T_2* of vessels, probably due to higher frequency information only being

sampled once in Partial Fourier imaging. This may be important with tumours as well, as they will require more higher frequency information compared to the liver parenchyma.

Compressed sensing is shown to be a promising acceleration technique, with high SNR images being produced in a very short length of time. However, considering the findings in Chapter 5 and 6, it would be advisable to have further studies comparing results with and without compressed sensing, particularly in a population with early HCC. Additionally, Boyarko *et al.*[87] found T_2* values obtained from an acquisition with compressed sensing only had a moderate correlation to T_2* measured without compressed sensing, which was probably due to motion between calibration scan and acquisition. Therefore, the calibration scan should be done immediately before the acquisition.

From our optimisations, we decided on the final parameters for clinically viable sequences for the application of 3D T_2* mapping of HCC. Further validation is required, particularly in a patient population at-risk of/with HCC.

A limitation of this work was that I did not measure SNR as measuring SNR accurately is challenging[96]. R^2 was deemed to be a reasonable surrogate for SNR. To avoid the need to assess SNR, ideally the sequence would be tested directly against the requirements, and this is the subject of later chapters of this thesis. Another limitation of this work is that assessment was only made in one volunteer for each change: ideally more volunteers would have been tested to verify the findings. However, this would have required more participants and would have been time consuming. Additionally, this optimisation process was performed on healthy volunteers and not on patients at-risk of/with HCC, where the results may have been different. However, it may not be feasible to undertake the optimisation process in that patient population. In subsequent chapters in this thesis the sequence was assessed in a phantom with concentrations to mimic HCCs and in patients with HCCs. A further limitation is that the requirements such as being able to detect

specific amounts of iron, were not assessed against biopsy paired iron quantification. However, it would be difficult to ethically justify taking biopsies.

7.5. Summary

This chapter outlined the process of optimisation of 3D ME GRE sequences for detecting iron sparing on both Philips and Siemens scanners, at 1.5 T and 3 T and gave the details of the parameters of the final sequences. Further verification that the sequences meet the requirements for detecting iron sparing in HCC patients is required, which is the subject of the next chapter.

8. Evaluation of optimised R_2* sequences in healthy volunteers

8.1. Introduction

Following on from the optimisation process described in Chapter 7, the sequences developed needed to be assessed in detail against the requirements for detecting iron sparing. To this end I have carried out several projects. Firstly, I evaluated data from 15 healthy volunteers scanned on a 1.5 T Siemens scanner and a further 15 healthy volunteers scanned on a 3 T Siemens scanner. Though experiments such as these are inevitably team efforts, I recruited the volunteers, performed the scanning on the 1.5 T scanner, provided radiographer training on the 3 T scanner, and assisted for the majority of the scans. I also set up the processing pipeline for the data including producing the code to help with this. I supervised the data analysis by a graduate intern at Perspectum, and I performed the statistical analysis, and evaluated the results.

It was also important to confirm that the sequences on both the Siemens and Philips 3 T scanners gave the same results as both scanners were going to be used for scanning early HCC patients as part of the SELINA study. To this end we scanned 5 volunteers on both the Siemens 3 T scanner and the Philips 3 T scanner. I recruited and organised the scanning. I was present for the scanning on the 3 T Philips scanner, although I did not do the scanning myself. I did the majority of the scanning of the participants on the 3 T Siemens scanner. I did all image analysis and statistics in relation to this data.

8.2. Evaluation of sequences at 1.5 T and 3 T

8.2.1. Introduction

To evaluate whether the optimised sequences would improve detection of iron sparing in HCC patients, we scanned 15 healthy volunteers at both 3 T and 1.5 T to see if the

requirements for detecting iron sparing were met. Although using Partial Fourier was essential to meet the requirement for the breath-hold lengths to be 16 s or less I wanted to have further verification (following the findings in Chapter 7) that using Partial Fourier would not impair the use of these sequences for detecting iron sparing. As healthy volunteers are generally able to manage longer breath-holds than patients, I compared results with and without Partial Fourier. At 1.5 T I included both Partial Fourier 6/8 in the phase direction and Partial Fourier 7/8 in the slice and phase direction. At 3 T I only included Partial Fourier 6/8 in the phase direction as using Partial Fourier in the slice direction resulted in susceptibility issues (more evident at 3 T due to the higher field strength).

8.2.2. Methods

Participants

15 volunteers were recruited for scanning at 1.5 T, and a further 15 volunteers were recruited for scanning at 3 T. All participants gave their written, informed consent. All volunteers had no known liver disease. The scanning of participants on the 1.5 T scanner was done under the auspices of the Clinical Affairs department at Perspectum for technical development purposes. The 3 T scanning was performed as part of the NICOLA Trial (Ethics: 13/SC/0243), as part of the healthy volunteer cohort.

MRI Protocol

The scanners used were a Siemens 3 T Prisma scanner and a Siemens 1.5 T Aera scanner. The MRI protocols at both 3 T and 1.5 T included the optimised 3D sequences, given in Section 7.3.4. At 3 T there was one optimised sequence, with Partial Fourier 6/8 in the phase direction. For 1.5 T the optimised sequence had Partial Fourier of 7/8 in both phase and slice direction. However, data with Partial Fourier 6/8 in the phase direction

only was collected for comparison, as both had the same breath-hold length, and each have different strengths and limitations, as covered in Chapter 7. Using Partial Fourier in the slice direction can result in T_2* shortening, therefore it might be preferable to use Partial Fourier only in the phase direction. However, if only using it in the phase direction a higher fraction is required to bring the breath-hold length down, and this could create worse delineation of higher frequency information in the image. The protocols also included the optimised 3D sequence with no Partial Fourier. The breath-hold length for these sequences would not be feasible in the patient population but allowed for assessment of the impact of Partial Fourier in the optimised sequences.

The robust 2D ME GRE, MOST, described in Section 3.7 was included to allow comparison with a "ground truth" R_2* in the central slice. Although not optimised for this specific application this sequence has high resolution, thin slice thickness, 8 echoes and final echo times of 19.06 ms and 9.84 ms for 3 T and 1.5 T respectively. Parameters for this sequence are given in Section 3.7.

At both field strengths, the optimised sequences were repeated immediately to check for repeatability. At 3 T an additional repetition was also performed at the end of the protocol (after some other sequences had been acquired), to see if this had any impact on the results. The summary of the sequences analysed in this work is given in Table 35.

Field Strength	Sequence	Slabs analysed	Repeats analysed
1.5 T	Optimised 3D sequence with Partial Fourier 7/8 in both Slice and Phase direction	2	1
	Optimised 3D sequence with Partial Fourier 6/8 in Phase direction	2	1
	Optimised 3D sequence with no Partial Fourier	2	1
	MOST	1 slice only	0
3 T	Optimised 3D sequence with Partial Fourier 6/8 in Phase direction	2	2 (final repeat at end of protocol)
	Optimised 3D sequence with no Partial Fourier	2	2 (final repeat at end of protocol)
	MOST	1 slice only	0

Table 35 Protocol of sequences for 1.5 T and 3 T scanning

Although most participants were scanned with the full protocol, there were some cases where data was partly missing. If fewer than 15 cases were included in statistical analysis this is mentioned in the results.

Image Analysis

The MAGO technique was used to generate T_2* maps[65]. A junior analyst (4 months' experience) performed image analysis, under my close supervision.

Firstly, analysis was performed on slices located at the centre of the liver. The MOST acquisition was positioned at the centre of the liver. For the optimised sequences, the central slice was typically selected as the centre of the second slab, which usually was also in a similar location to MOST. However, in some instances, due to liver shape and location of acquisition the central slice of the second slab was not appropriate so the analyst selected an alternative slice, generally by matching the MOST slice location.

Segmentation was performed, using a semi-automatic technique: a U-NET segmentation method [90] was used, followed by manual adjustment. Vessels were included as the impact on the median and IQR of pixels in the segmentation by including vessels

was deemed to be minimal (this was verified on multiple cases). Artifacts were not excluded.

Following the central slice analysis, a slice 24 mm below the dome of the liver was located, termed the upper slice. This was found using the upper slab or T_1-weighted VIBE sequence (if upper slab did not contain the dome of the liver). Segmentation was then performed on these slices using the same method as for the central slice.

Medians and interquartile range from the pixels within the segmentation were calculated. All T_2* values were converted to R_2*.

An example of the analysis (segmentation and ROIs) is given in Figure 48.

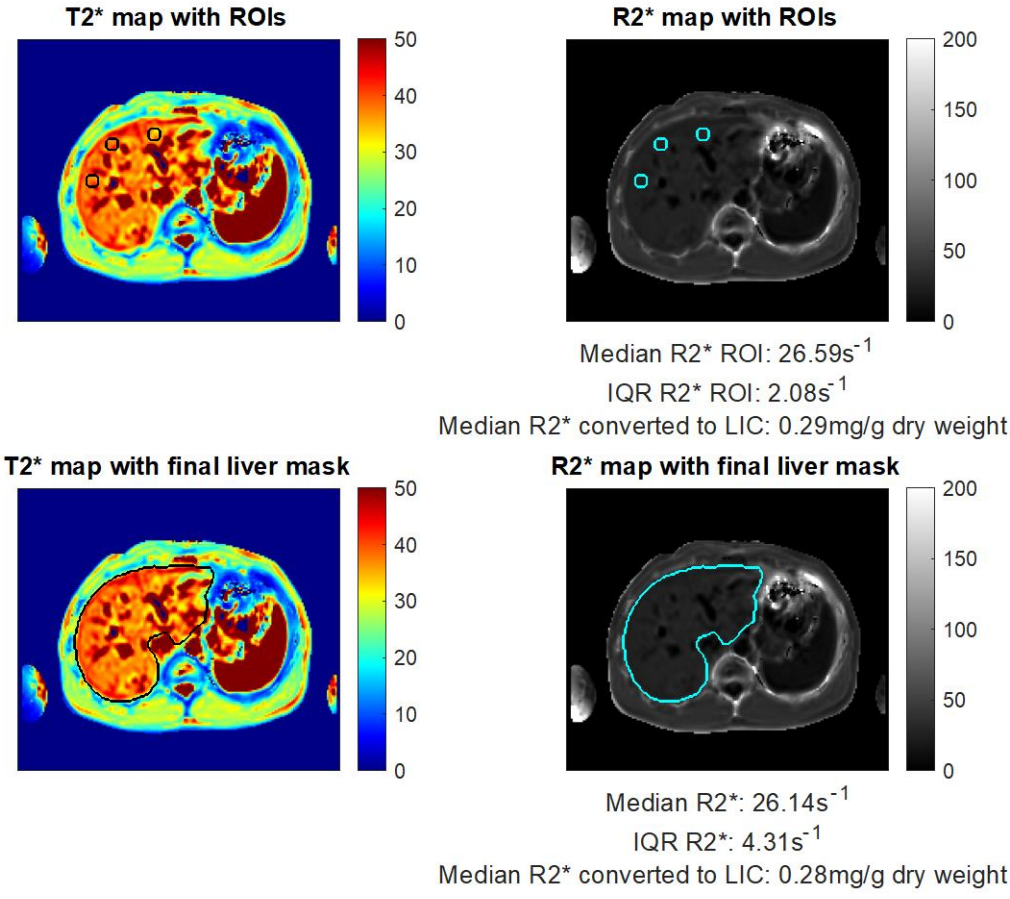


Figure 48 Central slice analysis for participant scanned at 1.5 T. These are the maps from the Partial Fourier 6/8 sequence.

Evaluation

Requirements for detecting iron sparing were presented in Section 3.8.4. To evaluate whether the sequences are likely to meet these requirements a number of experiments were performed. A brief summary of the assessments against each requirement is given in Table 36. The focus in this work was assessing against requirements 1 and 2. As well as performing these assessments on the optimised sequences with Partial Fourier, they were also performed on the sequences without Partial Fourier to see if using Partial Fourier made any difference to the results.

Requirement	Assessment
1. Measure iron concentration from 0 mg/g dw to 3 mg/g dw +/- 0.1 mg/g dw.	Compare median R_2* from segmentation placed on central slice of optimised acquisition with MOST acquisition ("ground truth"). Difference with MOST should be below 0.1 mg/g dw. Assess repeatability to check that the same measurements are obtained to within 0.1 mg/g dw, if not this indicates that the sequence is not robust enough for quantifying low iron levels.
2. Be able to detect differences of 0.2 mg/g dw or more of iron concentration within the liver (for tumour identification)	Check for homogeneity within and between slices (assuming iron is homogeneous within volunteer), by comparing median R_2* from segmentation between upper and central slice and by using inter-quartile range of R_2* segmentation. Ideally heterogeneity due to artifacts would be less than 0.1 mg/g dw to be able to assess 0.2 mg/g dw differences
3. Detection of tumours 1 cm or more	This is difficult to assess in healthy volunteers.
4. Breath-hold length of 16 s or less	Achieved: 1.5 T breath-hold length: 16 s (for both Partial Fourier variations) 3 T breath-hold length: 15 s
5. Full liver coverage	Achieved: liver coverage can be obtained using 3-4 x breath-holds. In this work only two upper slabs of the liver were acquired/analysed.
6. Needs to be deployable on standard clinical MRI scanners without modifying their software	Achieved.

Table 36 Requirements for detecting iron sparing, and how these were assessed

As indicated in Table 36, to assess against the first requirement, firstly, the median R_2* from the 3D acquisitions and MOST were compared. Most participants had relatively low R_2* values, indicating that their iron concentrations would be similar to (or even lower

than) the HCC population, so were well suited for this comparison. To compare results between MOST and the 3D acquisitions, scatter plots and Bland Altman plots were used, along with Spearman's rho (correlation) and ICC (agreement). Although ROIs were used for comparison with MOST in previous chapters, I decided to use the segmentation median R_2* as this should still be robust against extreme values and is also less affected by inter-operator variability, where the decision of where to place the ROIs can have a big impact on the results. Additionally, it saves time.

Another important aspect to assess is repeatability: it is important that the sequence gives comparable results in the same patient when repeated, so it can be reliably used for detection of HCC. Therefore, the segmentation R_2* median values of the central slice were compared between repeats of the acquisitions. For the 3 T data an additional repeat was acquired (after a gap in which other sequences were scanned), so that was also compared, to see if there was any impact on R_2* results due to increased difficulty breath-holding towards the end of scan for example. Like the MOST comparison, the results from the repetitions were compared using scatter and Bland Altman plots, along with Spearman's rho and ICC.

To assess Requirement 2, variation within the liver in the volunteers was assessed. In healthy volunteers it would be expected that iron distribution in the liver would be homogenous as they are unlikely to have any underlying liver disease which may result in heterogenous iron distribution[92]. Therefore, large amounts of variation are likely to be due to artifacts, and if this is the case then the acquisition may not be suitable for detecting small changes in iron throughout the liver. To assess for homogeneity through-plane, the upper slice segmentation median R_2* was assessed against central slice median R_2*. Once again, scatter plots and Bland Altman plots were used, along with Spearman's rho and

ICC. To assess for variation in-plane within the liver, the IQR of the segmentations of upper and central slice was calculated. These were plotted using a box plot.

If the results given from the statistical analysis showed that some cases did not meet the requirements, then further investigation was made by looking at the individual cases and evaluating why they did not meet them.

Relevant iron concentration thresholds for Requirements 1 and 2 are iron differences of 0.1 mg/g dw and 0.2 mg/g dw. Using Hankins' iron conversion equation[58], at 1.5 T a difference of 0.1 mg/g dw in LIC corresponds to 3.57 s^{-1} and a difference of 0.2 mg/g dw corresponds to 7.14 s^{-1}. At 3 T, using Hankins'[58] and Alam's[60] equations, a difference of 0.1 mg/g dw corresponds to 7.33 s^{-1} and a difference of 0.2 mg/g dw corresponds to 14.7 s^{-1}. Ideally within the liver and between different sequences/repeats the differences in R_2* would be less than the 0.1 mg/g dw thresholds.

8.2.3. Results

For the patients scanned at 3 T, the MOST segmentation R_2* median (converted to LIC using Alam's[60] and Hankins' equation[58]) gave LIC for these participants ranging from 0.30 mg/g dw – 3.03 mg/g dw, with median of 0.44 mg/g dw, lower quartile 0.36 mg/g, upper quartile 0.59 mg/g dw. Although the participant with LIC of 3.03 mg/g dw was relatively far from the other values (2.36 mg/g dw from the next highest LIC) it was still included in all analysis as some patients with HCC can have mild iron overload[43] and was only just outside the requirement of measuring up to 3 mg/g dw (when rounded it is 3.0 mg/g dw).

For the patients scanned at 1.5 T, the MOST segmentation R_2* median (converted to LIC using Hankins' equation[58]) gave LIC in these participants ranging from 0.18 mg/g dw – 2.52 mg/g dw, with median of 0.38 mg/g dw, lower quartile 0.36 mg/g, upper

quartile 0.52 mg/g dw. As with the 3 T participant with the higher LIC, although the participant with LIC of 2.52 mg/g dw was relatively far from the other values (1.37 mg/g dw from the next highest LIC) it was still included in all analysis.

Figure 49 shows scatter plots and Bland Altman plots, comparing the median R_2* from the segmentation the central slice of the optimised sequences with the 2D MOST sequence. Table 37 gives the statistical results for the comparisons. The correlation and agreement were excellent for all sequences. The bias was below the +/- 0.1 mg/g dw threshold for all sequences, both at 3 T and 1.5 T. The LoA at both 3 T and 1.5 T were outside of the +/- 0.1 mg/g dw threshold, although for 1.5 T they were close to the threshold. The LoA were all less than differences equivalent to 0.2 mg/g dw threshold. There was no clear difference between the results with and without Partial Fourier.

The cases where the differences were greater than the 0.1 mg/g dw or 0.2 mg/g dw thresholds (and hence causing the wider LoA) often had higher R_2* i.e., higher LIC, perhaps indicating that the optimised sequence has limitations for accurate measurement of higher R_2*. Upon investigating further, the short T_2* decay may be the cause of the different results between the optimised sequence and MOST in high R_2* cases: it is difficult to obtain a good fit when the decay is faster. Additionally, there seems to be more noise in the signal. This is shown in the plot of the decay in the signal of a high LIC case at 3 T with No Partial Fourier, in Figure 50 part (ii), which shows variation in the measured signal from an exponentiation curve. As the model includes fat, the model's estimate of fat has resulted in the fitting of an oscillating curve to the signal. However, for the MOST acquisition, the model does not estimate an oscillating exponential curve. The MOST signal and curve indicates that no fat is present, and therefore the variation in the signal in the 3D acquisition is likely due to noise. Therefore, the MOST acquisition seems better, perhaps due to the echoes being closer together and the signal being less

noisy due to having repeats. A plot of the T_2* signal decay using MOST is shown in Figure 50 part (i).

There were a few cases which did not have high LIC and still had larger differences compared to MOST. One example at 3 T (with a positive difference to MOST of 5.8 s^{-1}/7.1 s^{-1} for No Partial Fourier/Partial Fourier 6/8 s^{-1}) looks to be potentially due to proximity to the bowel, which has impacted the optimised sequence but not MOST (despite being in the same location). Another example at 1.5 T also had T_2* shortening close to the bowel for Partial Fourier Phase and Slice direction 7/8, resulting in difference with MOST of 3.8 s^{-1}. This was not so noticeable on the Partial Fourier Phase 6/8 and no Partial Fourier slices, which also had a much smaller difference with MOST (<2 s^{-1}). MOST was not impacted by T_2* shortening but was at a different slice location (13.7 mm higher).

Figure 51 shows scatter plots and Bland Altman plots, comparing the median R_2* from the segmentations of the central slice from the initial acquisition of the optimised sequences with the repeat of the sequence for 3 T. For the 3 T data there was a second repeat. The plots at 1.5 T are shown in Figure 52. For Partial Fourier 6/8 at 3 T one case did not have a repeat, so there are 14 data points. Table 38 gives the statistical results for the comparisons. For all repeats there was strong correlation and agreement. The bias and LoA were within the 0.1 mg/g dw thresholds at both 3 T and 1.5 T. The results with and without Partial Fourier were similar. At 3 T, where there was a second repeat, the bias and LoA were slightly larger than the first repeat, but they were still very small. Given the strong agreement and small biases and LoA, no individual cases were evaluated in more detail.

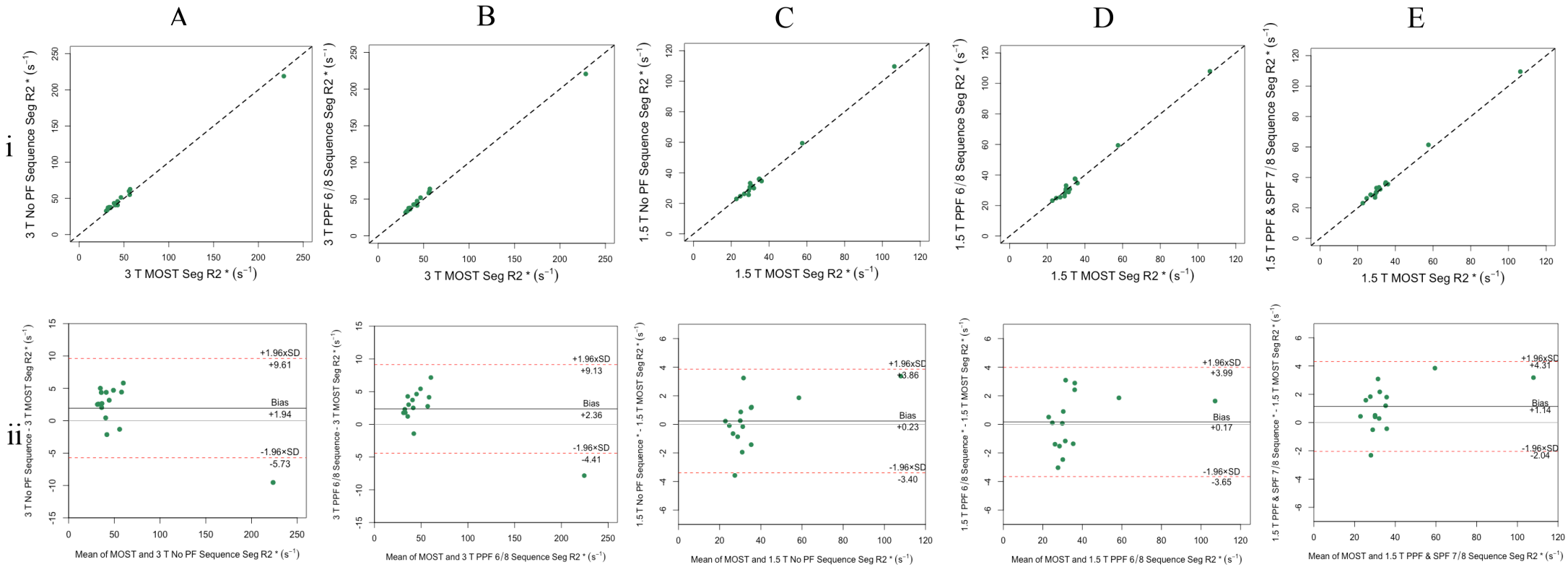


Figure 49 (i) Scatter plot and (ii) Bland Altman plot, comparing the median R_2* from the liver segmentations for optimised sequences with the median R2* from the liver segmentations for MOST. The variations optimised sequences compared with MOST are (A) 3 T No Partial Fourier, (B) 3 T Phase Partial Fourier 6/8, (C) 1.5 T No Partial Fourier, (D) 1.5 T Partial Fourier Phase 6/8 and (E) 1.5 T Partial Fourier Phase and Slice 7/8. The dashed line on the scatter plots is the identity line.

Field Strength	Sequences compared	Spearman's rho	Spearman's rho p value	ICC	ICC p value	Bland Altman bias (s^{-1})	Bland Altman lower limit of agreement (s^{-1})	Bland Altman upper limit of agreement (s^{-1})
3 T	MOST and No PF	0.95	p <0.001	1.00	p <0.001	1.94	-5.73	9.61
	MOST and Phase PF6/8	0.98	p <0.001	1.00	p <0.001	2.36	-4.41	9.13
1.5 T	MOST and No PF	0.94	p <0.001	1.00	p <0.001	0.23	-3.40	3.86
	MOST and Phase PF 6/8	0.95	p <0.001	1.00	p <0.001	0.17	-3.65	3.99
	MOST and Phase PF7/8, Slice PF7/8	0.96	p <0.001	1.00	p <0.001	1.14	-2.04	4.31

Table 37 Spearman's rho, intraclass correlation coefficient and Bland Altman statistics, comparing R_2* from segmentation of MOST with optimised acquisitions at 3 T and 1.5 T. Green indicates the value lies below the 0.1 mg/g dw LIC threshold, orange indicates it lies between 0.1 mg/g dw and 0.2 mg/g dw threshold, and red indicate it lies above the 0.2 mg/g dw LIC threshold.

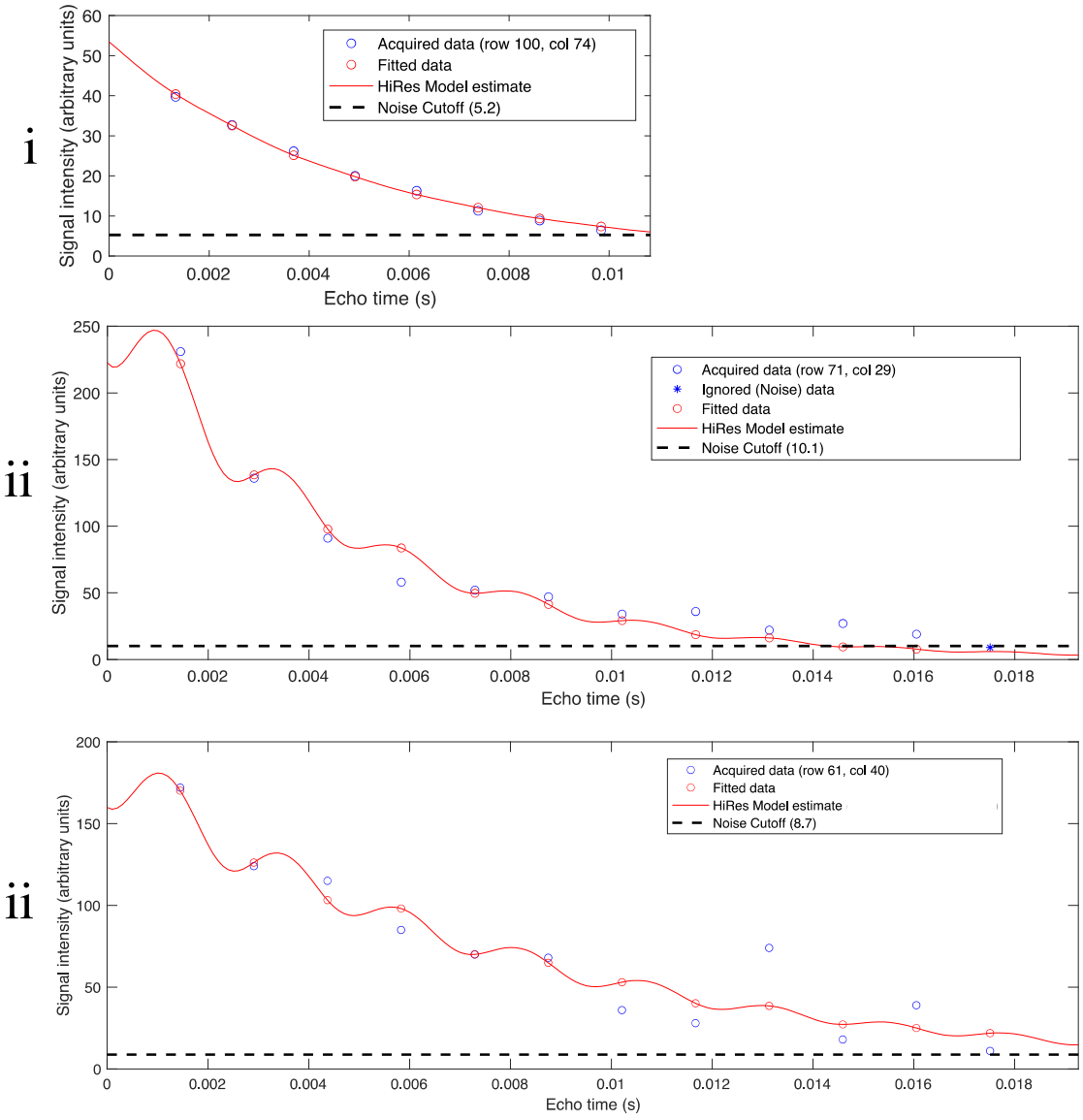


Figure 50 Plots of the signal decay in a pixel in the high LIC case at 3 T. (i) is a pixel from MOST, (ii) is a pixel from No Partial Fourier central slice and (iii) is a pixel from No Partial Fourier upper slice. The red points and line are the model estimate of the T_2* decay, and the blue points the acquired data. The dashed line indicated the noise floor, below which points are excluded (shown as stars).

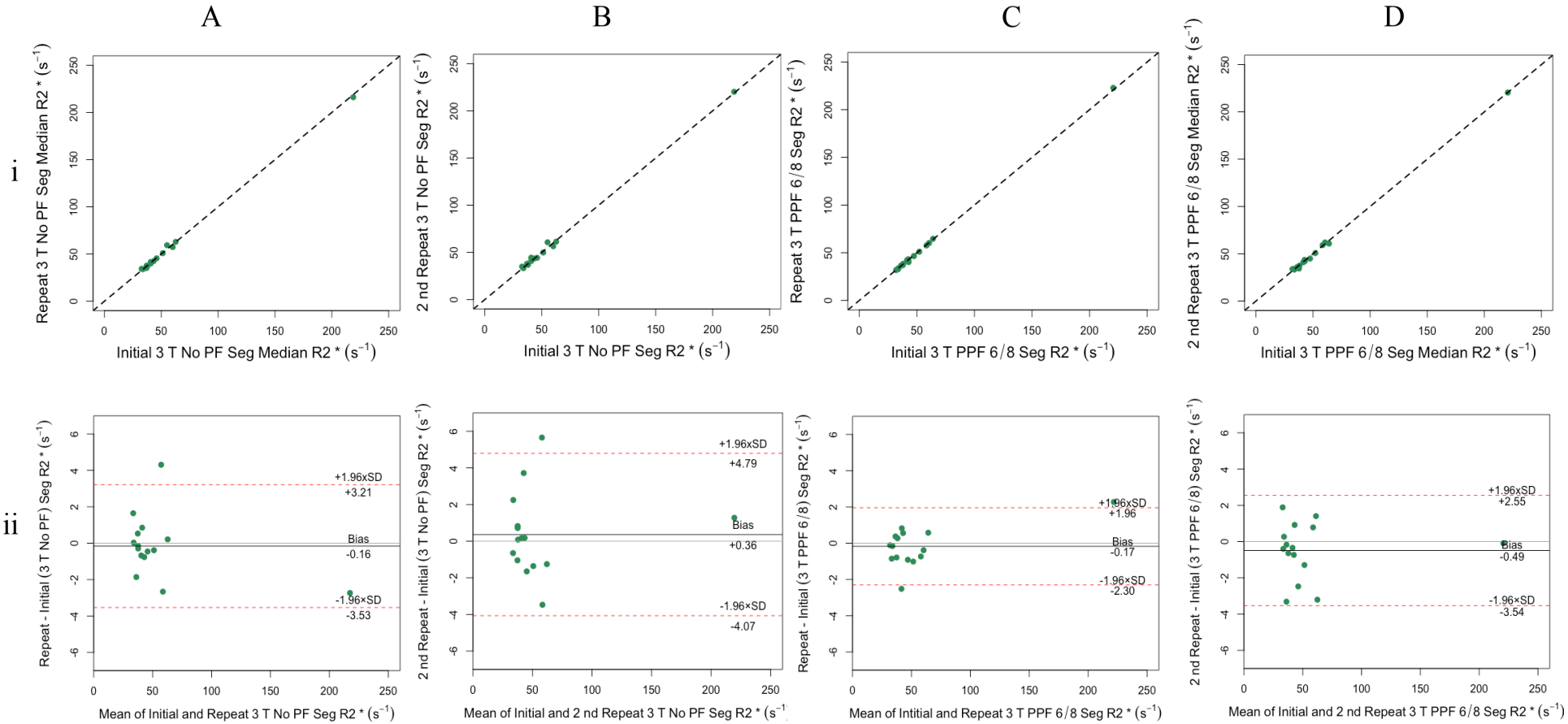


Figure 51 (i) Scatter plots and (ii) Bland Altman plots, comparing the median R_2* from the segmentation of the central slice for the initial and repeat acquisitions of 3 T optimised sequences. The variations optimised sequences are (A) 3 T No Partial Fourier, first repeat (B) 3 T No Partial Fourier, second repeat, (C) 3 T Partial Fourier Phase 6/8, first repeat, (D) 3 T Partial Fourier Phase 6/8, second repeat. The dashed line on the scatter plots is the identity line.

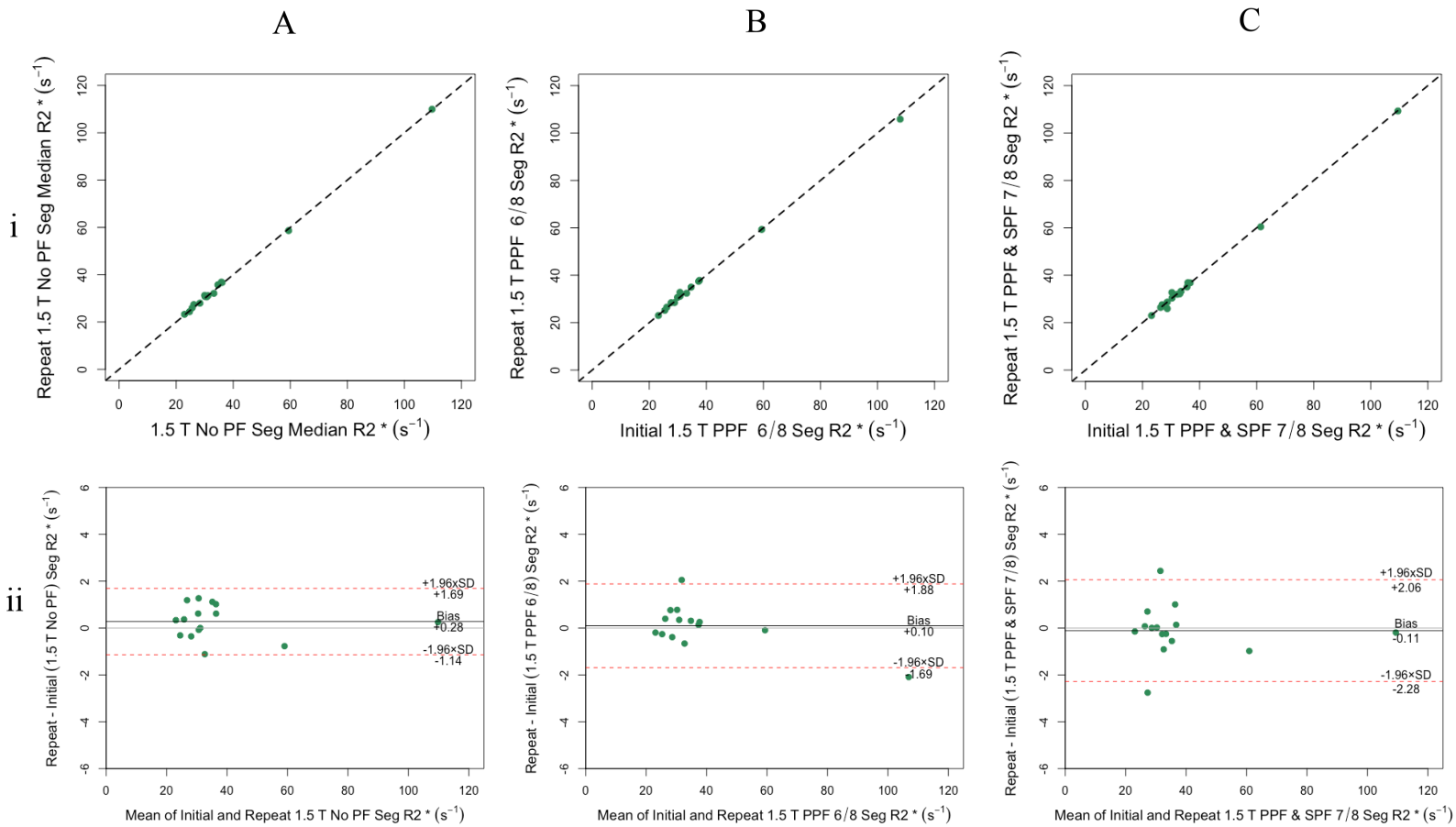


Figure 52 (i) Scatter plot and (ii) Bland Altman plot, comparing the median R_2* from the segmentation of the central slice for the initial and repeat acquisitions of 1.5 T optimised sequences. The variations optimised sequences are (A) 1.5 T No Partial Fourier, (B) 1.5 T Partial Fourier Phase 6/8, (C) 1.5 T Partial Fourier Phase and Slice 7/8. The dashed line on the scatter plots is the identity line.

Field Strength	Sequences compared	Spearman's rho	Spearman's rho p value	ICC	ICC p value	Bland Altman bias (s^{-1})	Bland Altman lower limit of agreement (s^{-1})	Bland Altman upper limit of agreement (s^{-1})
3 T	No PF Initial and Repeat 1	0.98	p <0.001	1.00	p <0.001	-0.16	-3.53	3.21
	No PF Initial and Repeat 2	0.94	p <0.001	1.00	p <0.001	0.36	-4.07	4.79 (
	Phase PF 6/8 Initial and Repeat 1	0.99	p <0.001	1.00	p <0.001	-0.17	-2.30	1.96
	Phase PF 6/8 Initial and Repeat 2	0.99	p <0.001	1.00	p <0.001	-0.49	-3.54	2.55
1.5 T	No PF Initial and Repeat	0.97	p <0.001	1.00	p <0.001	0.28	-1.14	1.69
	Phase PF7/8, Slice PF7/8, Initial and Repeat	0.95	p <0.001	1.00	p <0.001	-0.11	-2.28	2.06
	Phase PF 6/8 Initial and Repeat	0.99	p <0.001	1.00	p <0.001	0.10	-1.69	1.88

Table 38 Spearman's rho, intraclass correlation coefficient and Bland Altman statistics, comparing R_2* from repeats of the optimised acquisitions at 3 T and 1.5 T. Green indicates the value lies below the 0.1 mg/g dw LIC threshold, orange indicates it lies between 0.1 mg/g dw and 0.2 mg/g dw threshold, and red indicate it lies above the 0.2 mg/g dw LIC threshold

Figure 53 shows plots comparing the median R_2* from the segmentations of the central slice with the upper slice of the optimised sequences, Table 39 gives the corresponding statistics. At 3 T one case did not have an upper slice, so only 14 cases are included in the 3 T analysis. At both 1.5 T and 3 T there was strong correlation and agreement of central and upper slices across all sequences. At 3 T there was a small negative bias of the upper slice compared to the central slice, whereas the bias was positive for 1.5 T. At both field strengths the bias was less than the 0.1 mg/g dw threshold. At 3 T there were wide LoA, with the negative LoA being near to/larger than the 0.2 mg/g dw threshold. At 1.5 T the LoA were within the 0.2 mg/g dw threshold.

When investigating individual cases, most of the differences were near to, or less than the 0.1 mg/g dw thresholds, apart from the high LIC case at 3 T and the high LIC case at 1.5 T for Partial Fourier Phase and Slice direction 7/8 only. The high LIC case was the main reason why the 3 T LoA were so wide. The reason why the high LIC case has such a difference between central and upper slice is probably once again due to the difficulties of estimating T_2*, as mentioned previously in relation to the results comparing MOST and the optimised sequence. Figure 50 (ii) and (iii) shows the noisy signal and poor fit for both the central and upper slice of the high LIC case. For the high LIC case measured at 1.5 T using Partial Fourier 7/8 in phase and slice directions, the reason for the large positive difference between the upper and central slice is probably due to the Partial Fourier in the slice direction enhancing the T_2* shortening caused by the lungs. Apart from Partial Fourier 7/8 in phase and slice direction, at both 1.5 T and 3 T Partial Fourier and no Partial Fourier yielded very similar results.

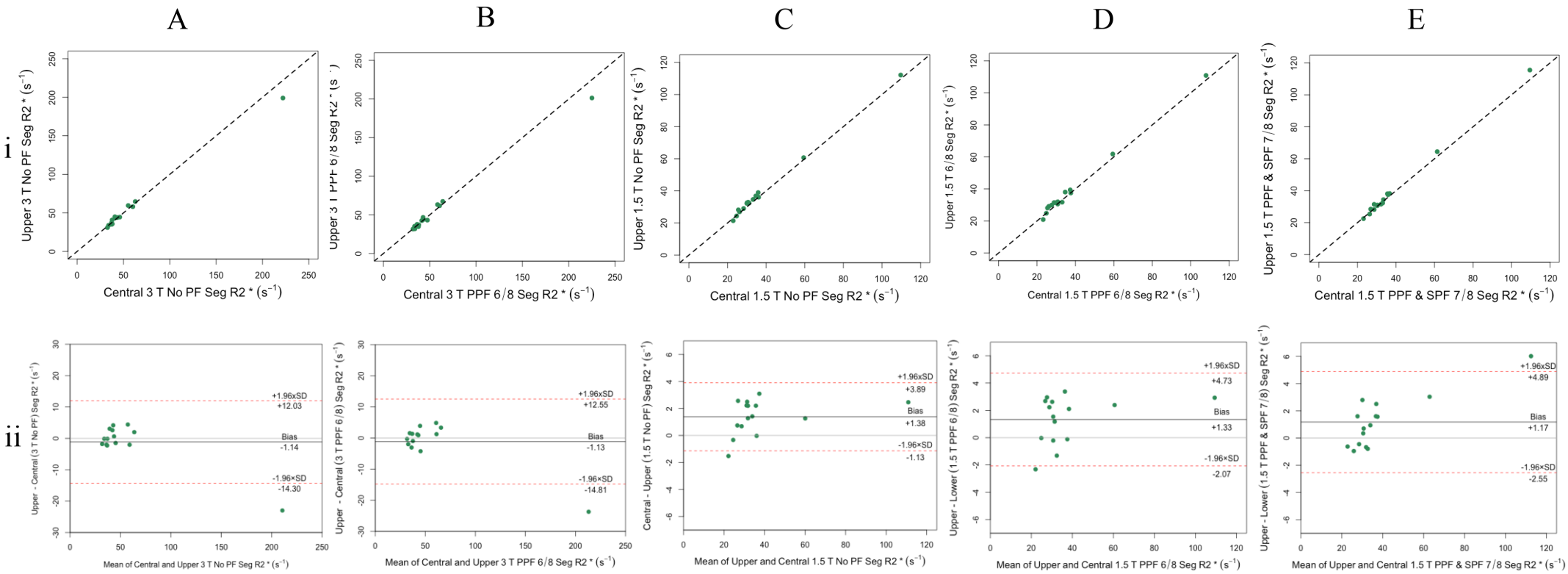


Figure 53 ((i) Scatter plots and (ii) Bland Altman plots, comparing the median R_2* from the upper and central slices for optimised sequences. The variations optimised sequences are (A) 3 T No Partial Fourier, (B) 3 T Phase Partial Fourier 6/8, (C) 1.5 T No Partial Fourier, (D) 1.5 T Partial Fourier Phase 6/8 and (E) 1.5 T Partial Fourier Phase and Slice 7/8. The dashed line on the scatter plots is the identity line.

Field Strength	Sequences compared	Spearman's rho	Spearman's rho p value	ICC	ICC p value	Bland Altman bias (s^{-1})	Bland Altman lower limit of agreement (s^{-1})	Bland Altman upper limit of agreement (s^{-1})
3 T	Central and Upper No PF	0.95	p <0.001	0.99	p <0.001	-1.14	-14.30	12.03
	Central and Upper Phase PF6/8	0.95	p <0.001	0.99	p <0.001	-1.13	-14.81	12.55
1.5 T	Central and Upper No PF	0.98	p <0.001	1.00	p <0.001	1.38	-1.13	3.89
	Central and Upper Phase PF7/8, Slice PF7/8	0.98	p <0.001	1.00 (	p <0.001	1.17	-2.55	4.89
	Central and Upper Phase PF 6/8	0.97	p <0.001	1.00	p <0.001	1.33	-2.07	4.73

Table 39 Spearman's rho, intraclass correlation coefficient and Bland Altman statistics, comparing R_2* of central and upper slices of the optimised acquisitions at 3 T and 1.5 T. Green indicates the value lies below the 0.1 mg/g dw LIC threshold, orange indicates it lies between 0.1 mg/g dw and 0.2 mg/g dw threshold, and red indicate it lies above the 0.2 mg/g dw LIC threshold.

Figure 54 contains the box plots showing the IQRs of the segmentation R_2* for central and upper slices. Table 40 gives the corresponding numerical values for the box plot. At 3 T one case did not have an upper slice, so the upper statistics for 3 T only have 14 data points. Otherwise, 15 cases have been included.

At both 3 T and 1.5 T the median IQR was greater than the 0.1 mg/g dw threshold, and the upper quartile of the IQRs was close to or above the 0.2 mg/g dw threshold. There were also some outliers with very high IQRs. The upper slices tended to have slightly higher IQRs compared to the central slice, and this was more pronounced at 1.5 T. For the central slice there was no clear trend of with/without Partial Fourier having higher IQR results compared to another, but for the upper slice at 1.5 T having Partial Fourier, and particularly Partial Fourier 7/8 in phase and slice direction gave higher IQR median and quartiles.

Upon investigation of the individual cases, the outliers with the highest IQR for 3 T and 1.5 T tended to be the cases with the highest LIC. For those cases with higher LIC, the reason for the larger IQR is probably once again due to the difficulty of estimating short T_2*. There were some cases which had high IQR and were not high LIC cases, and most frequently the reason for this was regions of T_2* shortening near the lungs. The reason for the higher IQRs for Phase and Slice Partial Fourier 7/8 upper slices (higher median and quartiles) looked to be mainly due to T_2* shortening near the lungs causing greater variations in T_2*, and hence R_2*.

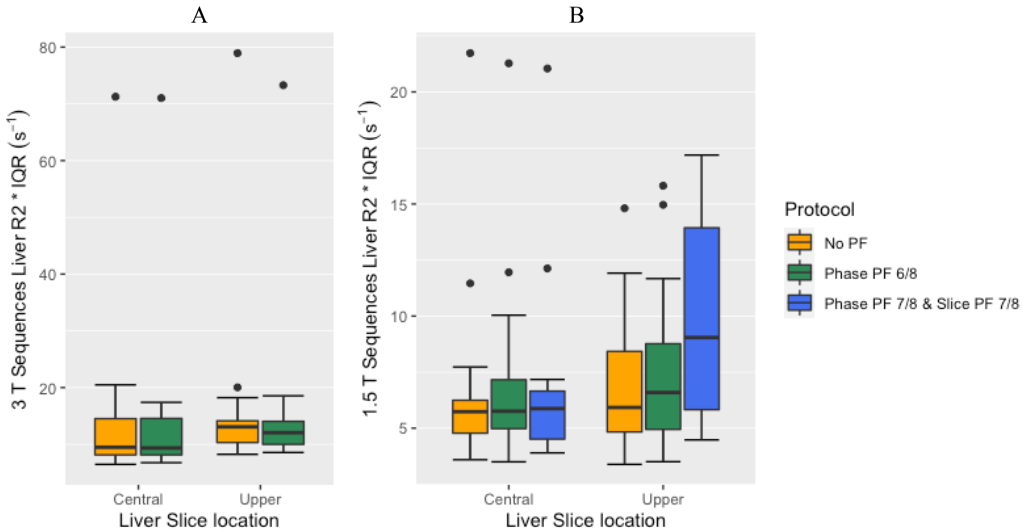


Figure 54 Box plots showing the R_2* interquartile range of the segmentation at (A) 3 T and (B) 1.5 T, for the variations of the optimised sequences

Field Strength	Sequences compared	Lower whisker (s^{-1})	Lower Quartile (s^{-1})	Median (s^{-1})	Upper quartile (s^{-1})	Upper whisker (s^{-1})	Outliers (s^{-1})
3 T	No PF Central	6.47	8.14	9.48	14.52	20.47	71.26
	Phase PF 6/8 Central	6.77	8.13	9.36	14.56	17.40	71.03
	No PF Upper	8.23	10.31	13.07	14.14	18.22	78.94, 20.03
	Phase PF 6/8 Upper	8.57	9.99	12.05	14.05	18.55	73.29
1.5 T	No PF Central	3.59	4.78	5.73	6.25	7.73	11.46, 21.73
	Phase PF 6/8 Central	3.50	4.98	5.76	7.16	10.04	11.95, 21.28
	Phase PF7/8 and Slice PF7/8 Central	3. 90	4.5	5.88	6.65	7.17	12.13, 21.04
	No PF Upper	3.39	4.83	5.92	8.43	11.92	14.81
	Phase PF 6/8 Upper	3.51	4.94	6.59	8.77	11.67	14.96, 15.81
	Phase PF7/8 and Slice PF7/8 Upper	4.48	5.83	9.05	13.94	17.18	No outliers

Table 40 Box plots statistics for the interquartile range of segmentation R_2* for Central and Upper slices. Green indicates the value lies below the 0.1 mg/g dw LIC threshold, orange indicates it lies between 0.1 mg/g and 0.2 mg/g threshold, and red indicate it lies above the 0.2 mg/g dw LIC threshold.

8.2.4. Discussion

This work aimed to assess whether the optimised sequences met the requirements for detecting iron sparing. Comparison with the "ground truth" 2D sequence and the repeatability both showed promising results: as there was a strong level of correlation and agreement, and the bias was less than 0.1 mg/g dw this indicates that the sequences can be used to accurately measure low iron concentrations, which is necessary for detecting iron sparing. Where differences between the ground truth and the optimised sequence was greatest was for highest LIC cases. This indicates that potentially an alternative sequence should be used for patients with LIC levels of more than 2.0 mg/g dw.

The strong agreement and correlation and small biases between central and upper slices indicate that the sequence could potentially be used throughout the liver. It was interesting that at 3 T the bias was negative (when comparing the upper to the central slice), whereas at 1.5 T the bias was positive. It would be expected that the bias would be positive, as we would predict T_2* to be shorter in the upper slices compared to the lower slices, due to through-plane susceptibility, and that this impact would be greater at 3 T. However, at 1.5 T the slice thickness was greater, and hence that might be the reason for the positive bias as the T_2* shortening would be worse due to the thicker slices. In the ELEGANCE study, with thicker slices, there was a larger positive bias found at 3 T when comparing higher and central slices. The results from this study therefore indicate that by reducing the slice thickness quality has improved in the upper slices. The LoA were wide at 3 T, but this was largely due to a case with higher LIC. This again indicates that the sequence may not be suitable for quantification of LIC throughout the liver at levels of LIC above 2 mg/g dw.

Many IQRs of the segmentations of the liver were larger than the requirement of 0.1 mg/g dw variation or less, although again, this was often in the participants with higher LIC. However, in the upper slices T_2* shortening near the lungs was still a cause for the wider IQRs. Therefore, it may be advisable to use a B_0 inhomogeneity correction to correct for the T_2* shortening to improve the quality of the upper slices[94].

There does not appear to be a clear trend suggesting that results using Partial Fourier differ significantly from using no Partial Fourier. This is reassuring as it indicates that the breath-hold can be reduced without causing a detriment in image quality. At 1.5 T, the IQR of the R_2* in the upper slice using Partial Fourier 7/8 in phase and slice direction was greater than with Partial Fourier 6/8 in the phase direction. This may be because the technique does not properly correct for the large phase shift that occurs due to the lungs in the slice direction. It is therefore probably preferable to use Partial Fourier 6/8 in the phase direction, as the acquisition will take the same amount of time.

A limitation of the study is that only healthy volunteers were imaged, so the liver tissue was different to the liver tissue in HCC patients who often have cirrhosis. However, given the LIC levels were low, which we would expect in most patients with HCC[17], it is still useful to assess the results in healthy volunteers, especially when it is easier for volunteers to spend time in the scanner compared to unwell patients.

Given that the population was healthy we did not perform any assessments which could help indicate as to whether we can detect spatially small (1-2 cm) variations in iron concentration, as associated with early tumours. A potential way to do this would be to look at delineation of vessels but they are quite different to tumours in terms of iron concentration and shape, so this was not deemed appropriate. Instead, a phantom was developed, as discussed in the next section.

The assumption that variations within slices and between slices was caused by artifacts is another limitation: ideally the slices would be assessed visually for artifacts, but this would be time consuming. However, most of the volunteers had low LIC and given that iron heterogeneity generally happens in patients with iron accumulation, and that even then it commonly isn't visualised on MRI[92], it is reasonable to assume iron distribution would be homogeneous and any differences in iron likely to due to artifacts.

Another limitation of this work was that only two individual slices were assessed. Ideally the full liver would have been evaluated, but in the interests of time this was not done. Additionally, the slices were not always in identical locations between sequences. In the future it may be preferable to use registration techniques to ensure slices in the same location are used.

We only used one vendor's scanner, Siemens. However, a Philips sequence was also optimised and, although we had fewer volunteers on that scanner, similar results were analysed as part of the cross-vendor study, see Section 8.3.

In conclusion the sequences optimised at both 3 T and 1.5 T can be used to quantify low iron concentrations. However, for iron concentrations of more than 2.0 mg/g dw, it may be advisable to use an alternative sequence that is optimised for those iron levels. Additionally, in some cases B_0 correction may be advisable in the upper liver to correct for T_2* shortening near the lungs.

8.3. Comparison across Philips and Siemens Scanners *in vivo*

8.3.1. Introduction

It is also important to assess the results between sequences *in vivo*, so that if different scanners are used in the same clinical trial, or for monitoring, there is confidence that they

give the same results. To this end five volunteers were scanned to assess the consistency of *in vivo* results.

8.3.2. Methods

Five volunteers were recruited. All volunteers gave their written, informed consent to be scanned. The Philips Ingenia 3 T scanning was performed under Nottingham University technical development research ethics (FHMS-36-1220). The Siemens Prisma 3 T scanning was performed under the NICOLA Trial ethics (Ethics: 13/SC/0243), as part of the healthy volunteer cohort.

The participants were scanned firstly on the Philips Ingenia 3 T scanner and then, within eight weeks of the Philips scan, on the Siemens Prisma 3 T scanner. The participants were scanned with the optimised sequences as given in Section 7.3.4.

The data were analysed by segmenting a central slice of the liver and then an upper and lower slice. The central slice was identified as the single slice used in other sequences in the protocol (not included in this study) and intersected the porta hepatis. The upper slice was located as the nearest slice to 24 mm below the dome of the liver, which was identified using a T_1-weighted scan. This location was used for consistency with the earlier study in this chapter. The lower slice was located, where possible, as the nearest slice to 24 mm above the peripheral slice of the lower liver, again identified using a T_1-weighted scan. However, in some cases the optimised sequence slabs did not cover the entire lower liver, so in the case that the location was either not included in the scan, the slice nearest 10 mm (non-Compressed SENSE) or 5 mm (Compressed SENSE) from the lowest slice of the slab was used instead (as quality can decline close to the periphery, and more so without Compressed sensing, hence the overlap between slabs which takes this into account).

Given the primary purpose of this work was the comparison of R_2* between different scanners, and not an assessment of quality, clear artifacts were excluded from the segmentation. However, a note was made when large amounts of the liver had to be excluded due to artifacts. A combination of a semi-automatic approach for segmentation (as described in previous *in vivo* studies in this thesis) and an entirely manual segmentation was used. An example of the central slice with segmentation is shown in Figure 55.

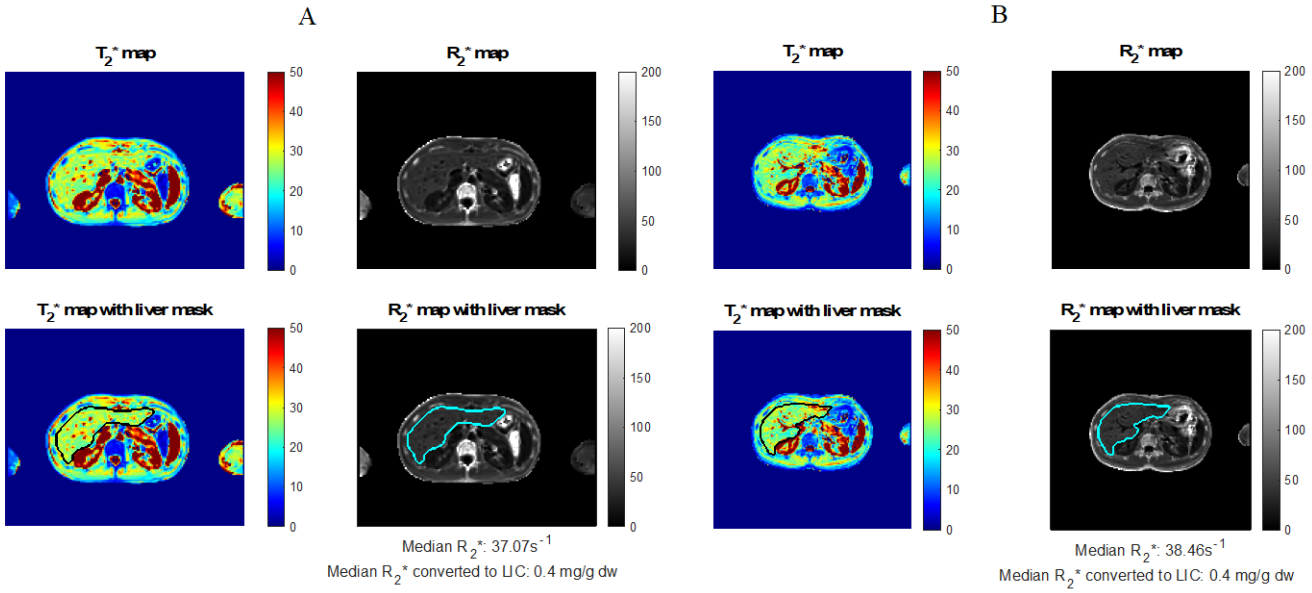


Figure 55 T_2* and R_2* maps with segmentation for central slice for Participant 4 for (A) Siemens and (B) Philips (without Compressed SENSE).

The results were recorded in tables for direct comparison. As in previous sections, statistical analysis included scatter plots, Spearman' rho (correlation), ICC (agreement) and Bland Altman plots (bias and LoA).

8.3.3. Results

Table 41 gives Spearman's rho and ICC for the volunteer comparisons between Siemens and Philips. Figure 56 show the scatter plots and Bland Altman plot comparing the results between the scanner for the central slice. Appendix C gives the plots comparing results from upper and lower slices. Table 42 gives the R_2* values for each segmentation and the equivalent LIC.

	Comparison	Spearman's rho (p value)	ICC (p value)
Central Slice	Siemens 3 T Optimised sequence and Philips 3 T Optimised Sequence with Compressed SENSE	1.00 (p = 0.016)	>0.99 (p <0.001)
	Siemens 3 T Optimised sequence and Philips 3 T Optimised Sequence with SENSE	1.00 (p = 0.016)	0.99 (p <0.001)
Upper slice	Siemens 3 T Optimised sequence and Philips 3 T Optimised Sequence with Compressed SENSE	0.70 (p=0.233)	0.99 (p<0.001)
	Siemens 3 T Optimised sequence and Philips 3 T Optimised Sequence with SENSE	0.70 (p=0.233)	0.99 (p<0.001)
Lower slice	Siemens 3 T Optimised sequence and Philips 3 T Optimised Sequence with Compressed SENSE	0.90 (p = 0.083)	0.99 (p <0.001)
	Siemens 3 T Optimised sequence and Philips 3 T Optimised Sequence with SENSE	0.90 (p = 0.083)	>0.99 (p <0.001)

Table 41 Spearman's rho and ICC for R_2* of Siemens compared to Philips sequences for five volunteers scanned on both scanners

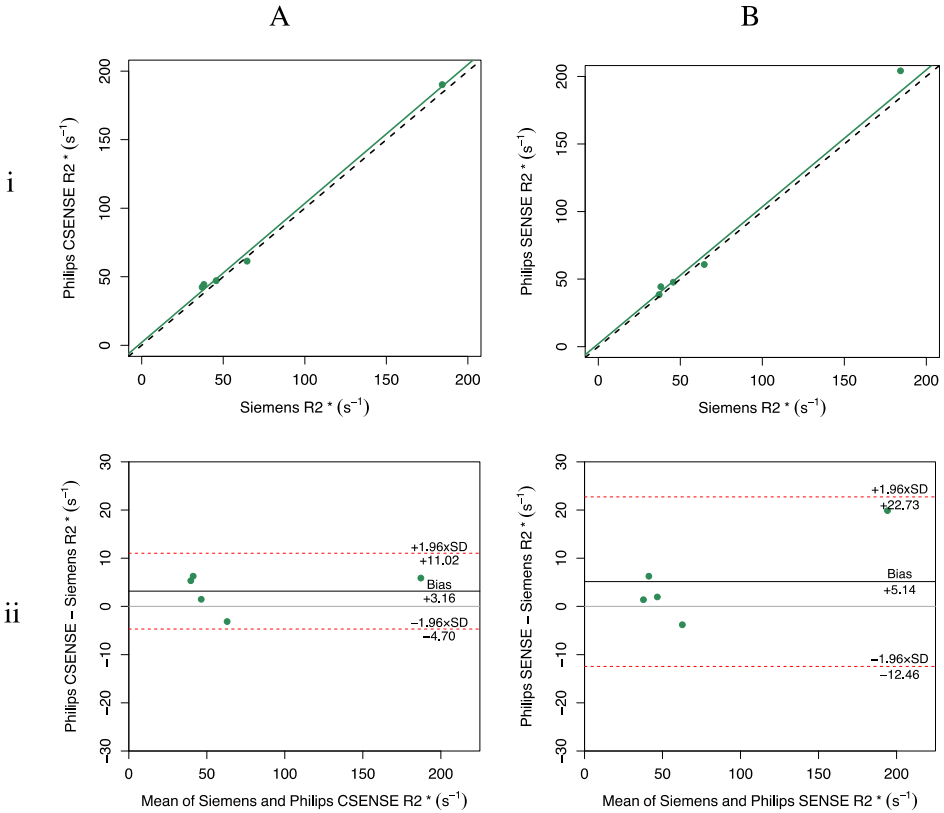


Figure 56 Volunteer central slice R_2*: (i) Scatter plots and (ii) Bland Altman plots comparing (A) the Philips optimised sequence with Compressed SENSE with the Siemens optimised sequence, (B) the Philips optimised sequence with SENSE with Siemens optimised sequence. On the scatter plot the solid green line is the line of best fit, the black dashed line is the identity line.

	Upper slice						Central Slice						Lower Slice					
Volunteer	Siemens R_2* (s^{-1})	Equivalent LIC (mg/g dw)	Philips Compressed SENSE R_2* (s^{-1})	Equivalent LIC (mg/g dw)	Philips SENSE R_2* (s^{-1})	Equivalent LIC (mg/g dw)	Siemens R_2* (s^{-1})	Equivalent LIC (mg/g dw)	Philips Compressed SENSE R_2* (s^{-1})	Equivalent LIC (mg/g dw)	Philips SENSE R_2* (s^{-1})	Equivalent LIC (mg/g dw)	Siemens R_2* (s^{-1})	Equivalent LIC (mg/g dw)	Philips Compressed SENSE R_2* (s^{-1})	Equivalent LIC (mg/g dw)	Philips SENSE R_2* (s^{-1})	Equivalent LIC (mg/g dw)
1	67.51	0.8	67.04	0.8	61.37	0.7	64.60	0.8	61.46	0.7	60.79	0.7	63.37	0.8	60.36	0.7	61.06	0.7
2	47.03	0.5	43.31	0.5	40.39	0.4	45.69	0.5	47.16	0.5	47.65	0.5	45.62	0.5	51.49	0.6	48.82	0.6
3	180.68	2.4	192.30	2.5	180.01	2.3	184.27	2.4	190.15	2.5	204.14	2.7	199.45	2.6	218.04	2.9	195.15	2.6
4	36.99	0.4	41.53	0.5	40.31	0.4	37.07	0.4	42.39	0.5	38.46	0.4	36.80	0.4	39.40	0.4	38.39	0.4
5	34.49	0.4	50.27	0.6	49.08	0.6	38.10	0.4	44.38	0.5	44.36	0.5	32.12	0.3	41.50	0.5	42.60	0.5

Table 42 R_2* values for the five volunteers at different slice locations, along with the iron concentration, calculated using Alam's[60] and Hankins'[58] equations

Table 41 shows that there was there was strong and significant correlation ($p<0.05$) for the central slice, however, for the upper and lower slice the correlation had $p>0.05$. There was strong and significant agreement for the upper, central, and lower slices. The Bland Altman plots show that the biases were always less than 7.33 s^{-1}. The bias was positive for slices at all slice levels, i.e., the Philips sequences gave higher R_2* compared to the Siemens acquisitions. Slice location did not have any clear impact on bias, neither did Compressed SENSE compared to SENSE. The LoA were wide (>7.33 s^{-1}).

Table 42 shows that for two of the participants, the largest difference in equivalent LIC between sequences was 0.1 mg/g dw. For another participant the largest difference was 0.2 mg/g dw. However, for the other two participants, Participant 3 and 5, the differences were larger: 0.6 mg/g dw and 0.3 mg/g dw respectively.

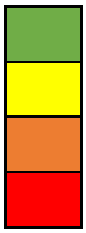
For Participant 3 there was no clear indication that one sequence gave lower/higher values than the others: for example, Philips SENSE gave the lowest R_2* for the upper slice, but the highest R_2* for the middle slice, compared to the other sequences. Given that this participant had higher LIC than the other participants, it is likely that the fitting is more difficult for the faster T_2* decay. This is particularly the case for these optimised sequences where the first TE is 1 ms/1.45 ms, which is when a lot of signal has already decayed for cases high in LIC. It may also be due to the scale of the T_2* colour map used for segmentation, which makes inhomogeneity due to artifacts less visible for patients with short T_2*. Therefore, the segmentation is more likely to include artifacts than for longer T_2* (lower LIC) cases.

For Participant 5 the Siemens sequence gave lower R_2* for all slices. This participant had difficulty holding their breath, which may have impacted the results (although severely artifacted areas were excluded).

The amount of liver excluded from the segmentation due to artifacts was also recorded, as shown in Table 43.

	Upper slice			Central slice			Lower slice		
Volunteer	Siemens % liver excluded	Philips Compressed SENSE % liver excluded	Philips SENSE % liver excluded	Siemens % liver excluded	Philips Compressed SENSE % liver excluded	Philips SENSE % liver excluded	Siemens % liver excluded	Philips Compressed SENSE % liver excluded	Philips SENSE % liver excluded
1	<12.5% of liver	0	0	More than 50% of liver	0	0	0	0	0
2	More than 50% of liver	12.5%-50% of liver	12.5%-50% of liver	<12.5% of liver	12.5%-50% of liver	12.5%-50% of liver	0	0	12.5%-50% of liver
3	0	0	0	0	0	0	0	0	0
4	12.5-50%	<12.5% of liver	0	0	0	0	0	0	0
5	<12.5% of liver	12.5%-50% of liver	0	<12.5% of liver	More than 50% of liver	0	0	0	0

Table 43 The percentage of the liver excluded from segmentation for each slice and sequence. The key to the colour coding is:



0 (no liver excluded

<12.5% of liver excluded

12.5%-50% of liver excluded

More than 50% of liver excluded

As can be seen from the Table, the lower slice was generally fully quantifiable (none of the liver had to be excluded from the segmentation). However, for the upper and central slices there were multiple instances were part of the liver had to excluded, and sometimes more than 50%. The Philips SENSE sequence had the least slices where liver had to excluded. Upon looking at the cases in more depth, the primary reasons for exclusion of liver were field inhomogeneity near the lungs/bowel and motion. Examples of two of the most impacted cases are shown in Figure 57 and Figure 58.

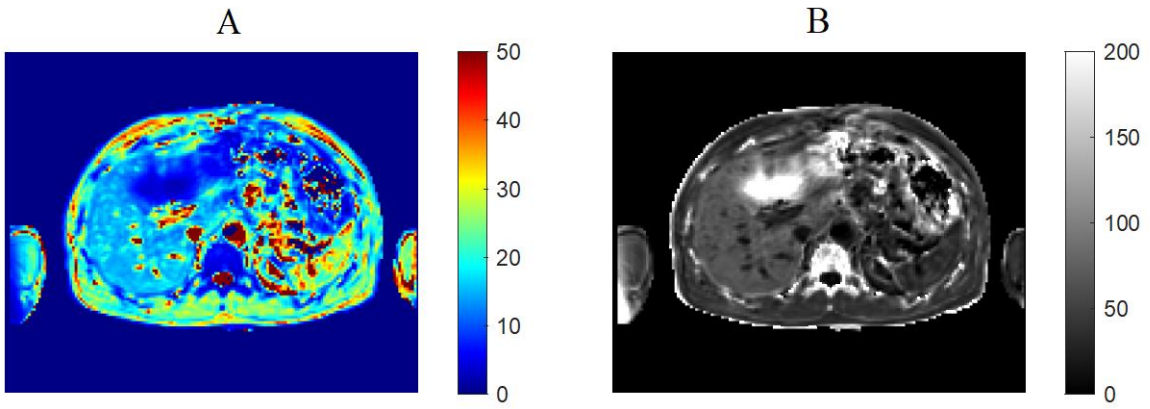


Figure 57 (A) T2* map and (B) R2* map for central slice of Participant 1 using Siemens protocol. T2* shortening looks likely due to proximity to the bowel.

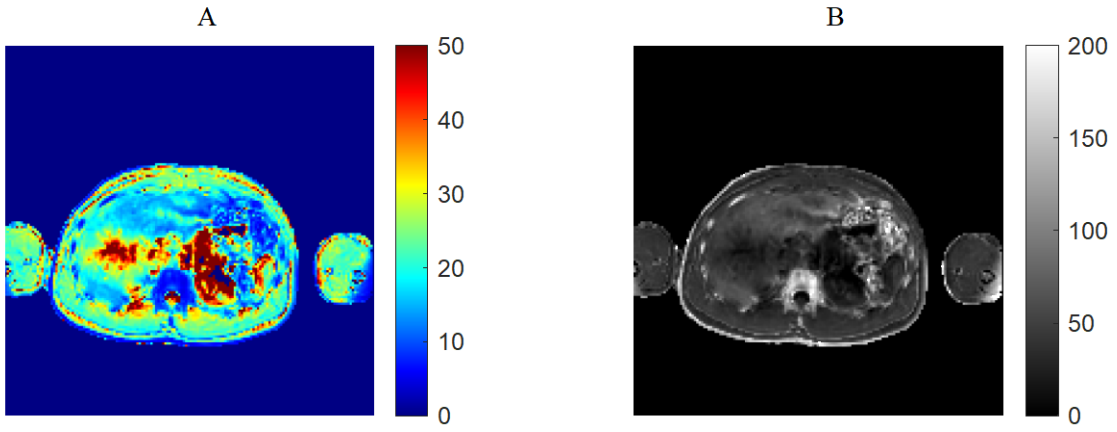


Figure 58 (A) T2* map and (B) R2* map for central slice of Participant 5 using Philips protocol with Compressed SENSE. Inhomogeneity in the liver looks likely due to motion.

8.3.4. Discussion

Consistency of R_2* measurements across manufacturers is important if patients are to be scanned on multiple scanners, for example during monitoring or in clinical trials. Although R_2* differences between liver and nodules are most important for detecting iron sparing, the R_2* values themselves may also have important implications e.g. given that iron declines over time in a nodule as it becomes cancerous[6] it may be useful to measure iron in a nodule over time and track this. Therefore, it was important to check consistency in R_2* between the two scanners where we had optimised the sequences. The strong agreement was reassuring, as were the small biases of less than 7.33 s^{-1} i.e., less than 0.1

mg/g dw difference between the sequences. However, the correlation was not significant in the upper and lower slices. Additionally, there were instances of large differences (>7.33 s^{-1}) in individual participants. One of these cases had LIC of over 2.0 mg/g dw. It may be that none of the results were accurate for the higher LIC case, as the first TE may have been too late to enable good fitting of the T_2* curve, as shown earlier in this chapter. Therefore, it may be recommended that an alternative sequence is used for participants with LIC of more than 2.0 mg/g dw.

Variation of 0.3 mg/g dw between results was also seen for Participant 5, which may have been due to the difficulties they had with breath-holding. For this participant large areas of the liver had to be excluded for some sequences and slices. This emphasises the need for good quality instructions and training for breath-holding.

Other cases also had to have large areas of the liver excluded, for example due to field inhomogeneity due to proximity to the bowel or lungs. This is consistent with the findings in Section 8.2 where, in 15 healthy volunteers scanned on the Siemens 3 T scanner, the large IQRs of segmentations (where artifacts were included) indicated inhomogeneity in R_2*. This indicates that it might be useful to include B_0 field correction, as reducing slice thickness further will increase the breath-hold length.

The results were consistent between slices, with no trend evident for higher R_2* at higher slices, consistent with the findings on the 15 volunteers scanned on the 3 T Siemens scanner, covered earlier in this chapter, and in contrast with the findings from ELEGANCE (Chapter 6). This is likely due to the thinner slices compared to in the ELEGANCE study.

Like the ELEGANCE study, quality was worse for Compressed SENSE vs SENSE, with more liver being excluded for Compressed SENSE. This indicates continued caution is needed when applying Compressed SENSE and that it may be preferable to use SENSE.

There have been some previous studies comparing T_2* measurements across vendors, which have shown good reproducibility across vendors[104,105], however these were at 1.5 T and using 2 D acquisitions. Although we reported R_2* and not T_2*, we also showed what would be classed as good repeatability for traditional LIC quantification, where the aim is to establish iron overload, for which the categories are broad (as mentioned previously)[56]. However, the reproducibility needs to be higher in the application of measuring iron sparing, and so even the results from the 2D 1.5 T may not be sufficient in this application. For example, one of the above mentioned studies showed inter-centre (i.e. vendor) reproducibility of 5.8% for liver T_2*[105].

More recently there has been a study comparing calibration curves of R_2* to LIC measured using R_2 across multiple sites and scanners[106]. This study involved 3 T scanners as well as 1.5 T, and the used of 3D sequences. They found that calibrations were highly similar across vendors[106], indicating reproducibility between scanners is likely to be high. However, again, this does not necessarily mean the reproducibility would be such to be of the level needed in the application of iron sparing. Given our small sample size, and the lack of other studies in this area, it would be advisable to do further cross-vendor comparisons. The small number of participants scanned at both sites reduces the power of the tests applied and our results may be affected by the small sample size.

Another limitation is that only 3 slices of the liver were analysed. A more comprehensive analysis of the entire liver is desirable, although more time consuming. Analysing three slices was a compromise between time limitations and covering more of the liver than one slice. Additionally, the slices were sometimes at different locations,

particularly when the full liver was not covered. It would be useful to repeat the study and ensure that the entire liver is covered by all sequences. Using image registration may help with identifying slices at the same location.

A further limitation is that these were healthy participants and so their liver tissue is different to patients with HCC. However, the iron levels were still representative of those which we would expect in most patients with HCC[17], and it would be less feasible to scan the same patients in both scanners at Oxford and Nottingham compared to healthy volunteers.

In conclusion, comparing the R_2* results across the two scanners showed that whilst they were in strong correlation and agreement, they might not be similar enough for the application of detecting iron sparing. However, changes such as improving breath-hold instructions and modifying the sequence for higher iron patients, may improve the results.

8.4. Summary

The results from the first study in this chapter indicate that the sequence largely meets the requirements which can be assessed in healthy volunteers, particularly in low iron cases. The second study showed that the results were also similar across vendors, although a different sequence for high iron patients and better training on breath-holding may help improve consistency across vendors

9. Evaluation of optimised R_2* sequences in "tumours"

9.1. Introduction

Scanning the healthy volunteers enabled testing some of, but not all, the requirements: in particular, detecting changes in iron associated with tumours and detection of tumours of 1 cm or more. Therefore, I designed and made a phantom to mimic tumours, which was scanned on both the 3 T Siemens and 3 T Philips scanners. It was not scanned on the 1.5 T Siemens scanner as the R_2* values of the phantom were not appropriate for that field strength. I played a key role in the design of the phantom, including giving input on the structure. I decided on the concentrations of the solutions and oversaw the mixing of the solutions in the laboratory. I filled the phantom spheres. The scanning on the 3 T Philips scanner was performed by my collaborators at University of Nottingham. I did the scanning on the 3 T Siemens scanner. I carried out all the image analysis and statistical processing.

Following the phantom scanning, patient recruitment began, which enabled a proof-of-concept type analysis: that this could be used in patients and could detect differences in R_2* in tumour compared to the liver. Scanning was performed by the radiographers/a senior physicist, and I assisted with some of the scans. I performed all image analysis. The script to combine the slabs into one volume was primarily developed by a colleague at Perspectum, but I oversaw the writing of it, suggested changes and implemented it.

9.2. Evaluation of sequences in a Phantom

9.2.1. Introduction

Testing the protocols that we had developed by scanning healthy volunteers, although helpful, had limitations for assessment against the requirements for a R_2* mapping sequence for detecting iron sparing. In particular, the ability to detect 0.2 mg/g dw

differences and tumours of 1cm or more is not possible to test in healthy volunteers. I did look at heterogeneity (non-physiological) within the healthy volunteers to assess whether, in theory, physiological differences of R_2* due to 0.2 mg/g dw differences in iron concentration could be detected. However, although this provides an indication of the ability of the sequence to meet this requirement, further evidence is desirable. To this end, we therefore designed and developed a phantom to mimic different iron concentrations in tumours and thus enable the assessment of three of the requirements for an iron detection R_2* mapping sequence:

1. Measure iron concentration from 0 mg/g dw to 3 mg/g dw +/- 0.1 mg/g dw.
2. Be able to detect differences of 0.2 mg/g dw or more of iron concentration within the liver (for tumour identification)
3. Detection of tumours 1 cm or more

9.2.2. Methods

Phantom Construction

The phantom structure consisted of a large 160 mm acrylic hollow outer sphere in which were positioned three plates with 20 mm, 15 mm, and 10 mm diameter hollow plastic spheres secured in place using epoxy. On each of the upper and lower plates there were five 20 mm and 15 mm spheres and four 10 mm spheres. On the central plate there were eight 20 mm, 15 mm, and 10 mm spheres. A photo of the phantom, along with diagrams of each of the layers is shown in Figure 59.

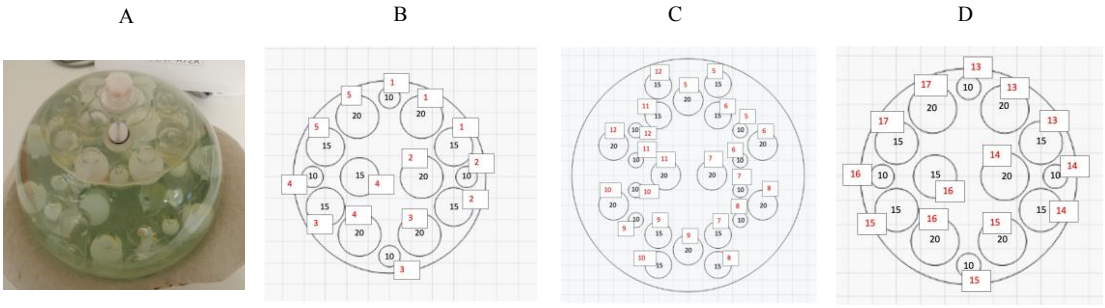


Figure 59 (A) Photograph of phantom, (B) Upper layer diagram, (C) Middle layer diagram, (D) Lower layer diagram. Black numbers inside the circles indicate the diameter (mm), red numbers in the black squares indicate the solution number (1-17)

The decision was made to use manganese (II) chloride ($MnCl_2$) to modulate R_2* and not iron. Recently several phantoms have been designed and constructed which use iron microspheres[84,86], however it would not be possible to use iron microspheres in this phantom as they are non-soluble and so need to be suspended in agar solution[84]. Filling the small spheres with agar solution with suspended microspheres in would be challenging. Additionally, I decided not to mimic any other substances in the liver and tumour aside from iron, due to the associated challenges. For example, I could have decided to include oil to represent fat in the liver and tumour, as fat alters R_2*[85] . Non-alcoholic steatohepatitis (NASH)) can lead to HCC[107] and a study found that 36.4% of the 1.1 cm – 1.5 cm HCCs studied contained fat, therefore it would be useful to include fat within the phantom. However, this would be challenging in the spheres, as peanut oil mixed with agar is typically used in phantoms to mimic fat in the liver[84].

$MnCl_2$ also changes R_1 so and nickel (II) chloride ($NiCl_2$) was used to modulate T_1, with the aim to keep R_1 as consistent as possible between the spheres ($NiCl_2$) changes R_2* as well, so this was taken into account when calculating concentrations). Although the aim was to keep R_1 consistent, we did not aim to mimic R_1 in the liver and tumours. In the future it may be preferable to mimic tumour and liver R_1. Target R_2* for the solutions were decided based on iron concentrations which would be expected in HCC

tumours and surrounding tissue[17]. 17 target iron concentrations 0.4 – 2.0 mg/g dw at 0.1 mg/g dw intervals were converted into R_2* using Hankins' and Alam's equations[58,60].

To work out the concentrations we can use the equations

$$R_1 = [MnCl_2]\, r_{1_{MnCl_2}} + [NiCl_2] r_{1_{NiCl_2}} + R_{1_{Water}}$$

$$R_2 = [MnCl_2]\, r_{2_{MnCl_2}} + [NiCl_2] r_{2_{NiCl_2}} + R_{2_{Water}}$$

R_1 and R_2 are the desired relaxation rates, $[MnCl_2]$ and $[NiCl_2]$ are the concentrations which we would like to find. $r_{1_{MnCl_2}}$, $r_{1_{NiCl_2}}$, $r_{2_{MnCl_2}}$ and $r_{2_{NiCl_2}}$ are the relaxivities of $MnCl_2$ and $NiCl_2$. $R_{1_{Water}}$ and $R_{2_{Water}}$ are the relaxation rates of pure water. Values for the relaxivities of $MnCl_2$ and $NiCl_2$ are available in the literature, as is the relaxation rate of pure water[108]. Using the known values in the equations above, they were solved for $MnCl_2$ and $NiCl_2$ for different desired R_2*. It was assumed $R_2 = R_2$* as these are freely diffusible agents. However, it should be noted shimming will impact R_2* but not R_2. It was aimed to keep R_1 constant at 9.8 s^{-1}, as this was the easiest number to achieve given the relaxivities of $MnCl_2$ and $NiCl_2$.

Two 300 ml stock solutions were made of $MnCl_2$ and $NiCl_2$ dissolved in pure water, from which small amounts were pipetted to make 50 cm^3 solutions of the desired concentrations, which included both $MnCl_2$ and $NiCl_2$ solutions, and additional water. These solutions were then syringed through small holes into the small spheres, which had already been secured onto the plates, and sealed with epoxy. The plates were then placed in the large sphere. In reality, tumour boundaries are very different to the boundaries of the small plastic spheres, particularly as the plastic walls were quite thick and bubbles could form on them. However, it would be complex and difficult to mimic tumour boundaries, so the decision was made to use plastic spheres. It may be useful to use spheres with thinner boundaries in future phantoms.

The remaining solutions were sealed into 25 mm plastic tubes. Given these tubes are larger, and so easier to image, the tubes were used to obtain reference values for the spheres. These tubes were placed in a cylindrical container 210 mm in diameter and surrounded by water.

A flood solution was also made to fill the large sphere. The desired R_2* was decided to be equivalent to 1 mg/g dw, as HCC patients can have background LIC of this value[17]. It was aimed for the R_1 of the flood solution to be the same as the spheres. A concentration amount of solution was made before being diluted to the desired concentrations, and with a large enough volume to fill the spheres. This was poured into the top of the large sphere, which was then sealed.

MRI scanning

The scanners used were the same scanners on which the optimisation was performed: a 3 T Philips Ingenia scanner and a 3 T Siemens Prisma (see Chapter 7). The optimised sequence parameters are given in Section 7.3.4. The tubes were only scanned on the 3 T Philips Ingenia. They were placed in the scanner and scanned with the MOST sequence, parameters for 3 T Philips scanners are given in Section 3.7. The temperature was not taken for the initial scan but was for a subsequent scan.

The spherical phantom was scanned on both the Philips 3 T and Siemens 3 T scanner. The abdominal coils were used, to match the *in vivo* protocol. On the Philips scanner the phantom was scanned with the optimised sequence with Compressed SENSE and the optimised sequence with SENSE. On the Siemens scanner the phantom was scanned the one optimised sequence developed on that scanner. Parameters for the optimised sequence are given in Section 7.3.4. For each sequence, data was acquired at three locations, each one centred on a different plate. The temperature was taken before and

after scanning. MOST was also acquired: this was used only for a reference value of the background flood solution. Although temperature was measured when the phantom was scanned, the temperature did change during the phantom scan, and may have been different to when the reference tubes were scanned. Temperature can influence T_2*[108], so in the future it would be preferable to keep the phantom at a constant temperature.

Image Analysis

The MAGO technique was used to generate T_2* maps[65]. For the tubes, 15 mm diameter circular ROIs were placed, one on each tube. Median T_2* and R_2* was calculated. For the background reference value using MOST, a 10 mm diameter circular ROI was placed on the background of the central slice and median R_2* found.

For the spheres, 10 mm diameter circular ROIs were placed on the 20 mm spheres and also in the flood solution area. 7.5 mm diameter circular ROIs were placed on the 15 mm spheres. 5 mm diameter circular ROIs were placed on the 10 mm spheres. For the optimised sequences the number of pixels within each circular ROI corresponded to 9, 5 and 1 pixels for the 10 mm, 7.5 mm and 5 mm diameter ROIs respectively. The median R_2* was found for each ROI. R^2 maps were also viewed, to give an indication of how well the measured data fitted with the model estimate.

Statistical Analysis

The ground truth values obtained from the MOST scan were firstly compared with the desired values. They were then compared to the values from each of the 3D sequences using scatter plots, Bland Altman plots, Spearman's rho (correlation) and ICC (agreement). Philips and Siemens R_2* results were compared directly to each other using the same statistics.

Assessment against requirements

The requirements assessed with the phantom were:

1. Measure iron concentration from 0 mg/g dw to 3 mg/g dw +/- 0.1 mg/g dw.
2. Be able to detect differences of 0.2 mg/g dw or more of iron concentration within the liver (for tumour identification)
3. Detection of tumours 1 cm or more

For assessing whether Requirement 1 had been met the MOST reference values were compared with the measured values using scatter plots, Spearman's rho (correlation), ICC (agreement) and Bland Altman plots. To specifically check that the results were equivalent to +/- 0.1 mg/g dw of MOST reference values all results were converted to iron concentration and compared directly to the MOST reference values converted to iron. At 3 T 0.1 mg/g dw difference corresponds to 7.33 s^{-1}, using Alam's[60] and Hankins'[58] equations

For assessing whether Requirement 2 had been met, concentrations with reference R_2* near to the reference background R_2* were compared with the background values to see if there was a difference between the sphere value and the background, and there was no overlap in the range/IQR (i.e., the error) of the sphere value and the background value. This was done using box plots: if the box plots did not overlap this shows that a difference could be detected. Given the small size of the ROIs this was only done for the large spheres, so the IQR values were meaningful, as the number of pixels within the ROI for the medium and small spheres was very small. As well as the box plots, the differences in the measured values of spheres vs background were compared with the reference differences of the concentrations with the background, for those concentrations close to the background values, for all sphere sizes.

Whether the size of the spheres had an impact on the results was evaluated in order to assess if Requirement 3 had been met.

9.2.3. Results

Firstly, reference values for the spheres were established by scanning the tubes using the MOST sequence on the Philips scanner. The scatter plot and Bland Altman plots comparing the results from the MOST sequence of the tubes with the desired values is shown in Figure 60. Spearman's rho was 1.00 ($p < 0.001$) and ICC was >0.99 ($p < 0.001$). There was a small negative bias, as seen in the Bland Altman plot. This could have been due to differences in the acquisition compared to those used to obtain the relaxivities of $MnCl_2$ and $NiCl_2$, as well as differences in shimming. The equivalent iron concentration of each of the measured R_2* values were the same as the desired equivalent iron concentration, apart from for concentration 12, which was different by 0.1 mg/g dw. The desired R_2* and equivalent iron concentration, along with measured R_2* and equivalent iron concentration, is given in the Appendix D.

The MOST scan was also repeated 8 weeks later. When comparing the R_2* from the repeated scan and the initial scan, Spearman's rho was >0.99 ($p < 0.001$) and ICC was >0.99 ($p < 0.001$) and bias +0.71 s^{-1}. More details of the repeated scan, including scatter plot and Bland-Altman plots comparing the initial and repeated results are given in Appendix D.

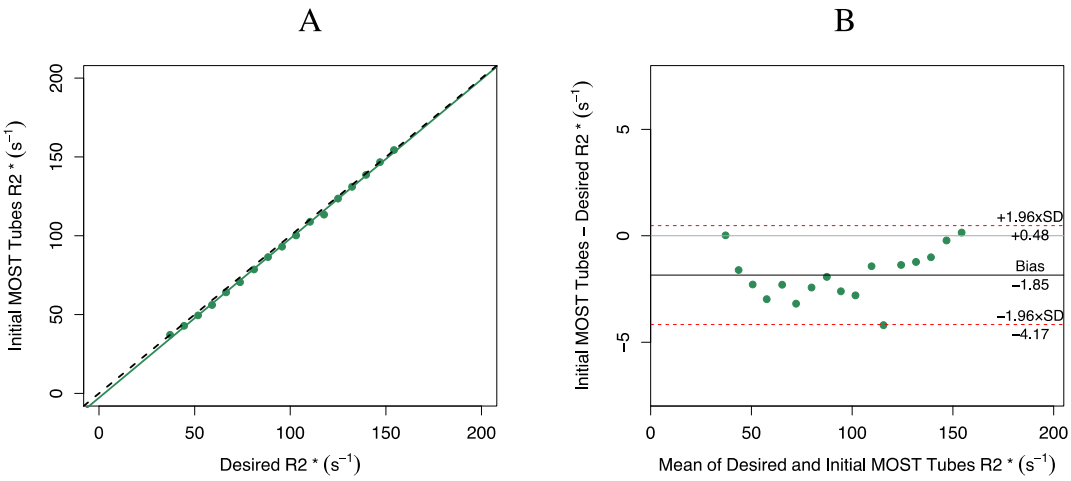


Figure 60 (A) Scatter plot and (B) Bland Altman plot comparing desired R_2* with measured R_2* for the MOST acquisition of the tubes. On the scatter plot the solid green line is the line of best fit, the black dashed line is the identity line.

The initial MOST results were used as reference, as these had the complete set of tubes (the repeated scan had one tube missing, see Appendix D for more details) and had similar results to the repeat. For the reference background, an ROI placed on the central plate of the spherical phantom scanned using MOST on the Philips scanner was used. This gave an R_2* of 73.18 s^{-1}, equivalent to 0.9 mg/g dw iron concentration. The desired R_2* was 81.07 s^{-1} equivalent to 1.0 mg/g dw of iron concentration.

The spherical phantom was scanned firstly on the Philips scanner. The temperature was recorded as 23.5 °C at the start and 20.6 °C at the end of the scanning. It was then transported to the Siemens scanner site. It was scanned 17 days after the Philips scan on the Siemens scanner. The temperature was recorded as 22.9 °C at the start and 23.6 °C at the end of the scanning.

Figure 61 shows the T_2* map, R_2* map and R^2 map for the middle plate for both the Philips and Siemens acquisitions. R^2 was generally 0.95 or more in the spheres for Philips but for Siemens R^2 was <0.95 in and around some of the spheres, and this was also the

case for some of the spheres on the other plates. The medium spheres with concentration 16 and 17 were excluded from both Philips and Siemens analysis as it seems there was air in sphere 16, shortening the T_2* of both 16 and 17. For the small spheres concentration 7 was excluded for the same reason. There also was no sphere with concentration 17 for the small spheres. Additionally, for the Siemens data only, the large sphere with concentration 10 had a very short T_2* so was also excluded. During transit some of the spheres had become loose, so large spheres with concentrations 7 and 11 were excluded from the Siemens analysis. Small spheres with concentrations 5 to 8 on the central plate also were not seen on any acquired images on the Siemens scan, indicating they probably had come loose and floated away from the plate. Given the above, the number of measurements to evaluate were less for Siemens, which lessens the power of any statistical calculations.

Table 44 gives Spearman's rho and ICC, and bias and 95% LoA from the Bland Altman plots comparing the median R_2* of the ROIs with the reference values. The corresponding scatter plots and Bland Altman plots comparing the measured values with the reference values are given in Appendix D. Table 45 shows the ROI median R_2* values and the equivalent iron concentration, along with the reference.

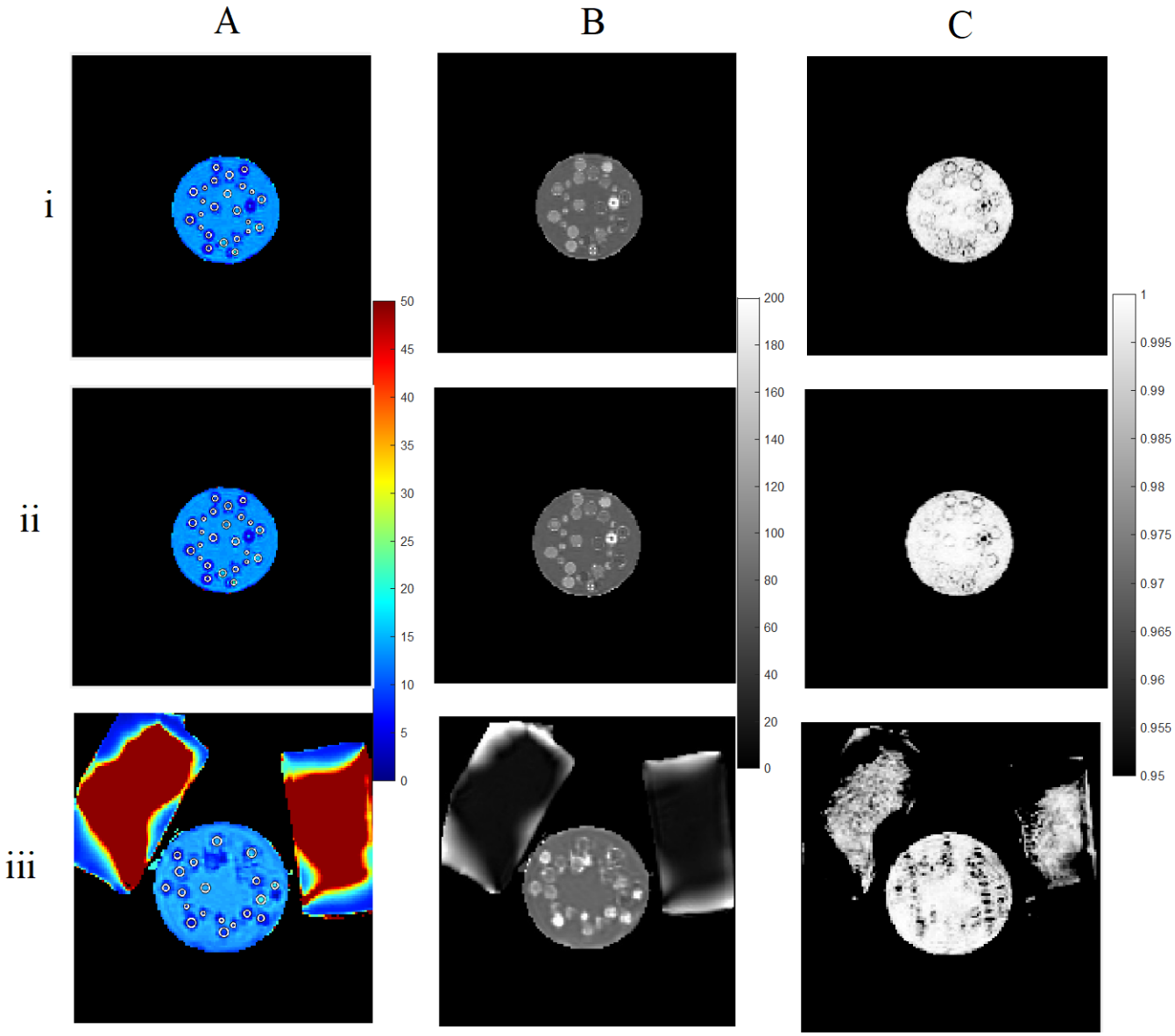


Figure 61 (A) T_2* map, (B) R_2* map and (C) R^2 map for (i) Philips Compressed SENSE, (ii) Philips SENSE and (iii) Siemens acquisitions of the middle plate. ROIs are shown on the T_2* maps.

Size of spheres	Scanner	Sequence	Spearman's rho (p value)	ICC (p value)	Bias (s^{-1})	Lower 95% limit of agreement (s^{-1})	Upper 95% limit of agreement (s^{-1})
20 mm	Philips	Modified sequence with Compressed SENSE	>0.99 (p <0.001)	0.99 (p <0.001)	-2.00	-11.79	+7.78
		Modified sequence with SENSE	0.99 (p <0.001)	0.99 (p <0.001)	-2.41	-12.81	+8.00
	Siemens	Modified sequence	1.00 (p <0.001)	0.99 (p <0.001)	-1.27	-14.42	+11.88
15 mm	Philips	Modified sequence with Compressed SENSE	0.92 (p <0.001)	0.84 (p <0.001)	+12.59	-22.40	+47.59
		Modified sequence with SENSE	0.91 (p <0.001)	0.88 (p <0.001)	+9.53	-21.66	+40.73
	Siemens	Modified sequence	0.88 (p <0.001)	0.73 (p = 0.005)	+18.84	-28.70	+66.37
10 mm	Philips	Modified sequence with Compressed SENSE	0.91 (p <0.001)	0.88 (p <0.001)	+12.39	-18.22	+43.00
		Modified sequence with SENSE	0.87 (p <0.001)	0.85 (p <0.001)	+12.82	-25.20	+50.84
	Siemens	Modified sequence	0.59 (p=0.049)	0.67 (p = 0.041)	+15.85	-44.98	+76.68

Table 44 Spearman's rho and ICC for the R_2* reference values compared with the measurements from the Philips and Siemens optimised sequences

		Large Spheres						Medium Spheres						Small spheres					
Reference R_2^* (s^{-1})	Equivalent iron (mg/g dw)	Philips Compressed SENSE R_2^* (s^{-1})	Equivalent iron (mg/g dw)	Philips SENSE R_2^* (s^{-1})	Equivalent iron (mg/g dw)	Siemens R_2^* (s^{-1})	Equivalent iron (mg/g dw)	Compressed SENSE R_2^* (s^{-1})	Equivalent iron (mg/g dw)	SENSE R_2^* (s^{-1})	Equivalent iron (mg/g dw)	Siemens R_2^* (s^{-1})	Equivalent iron (mg/g dw)	Compressed SENSE R_2^* (s^{-1})	Equivalent iron (mg/g dw)	SENSE R_2^* (s^{-1})	Equivalent iron (mg/g dw)	Siemens R_2^* (s^{-1})	Equivalent iron (mg/g dw)
37.17	0.4	37.12	0.4	36.57	0.4	33.37	0.3	40.60	0.4	40.42	0.4	41.94	0.5	44.26	0.5	43.82	0.5	130.38	1.7
42.86	0.5	41.80	0.5	41.69	0.5	37.56	0.4	41.03	0.5	39.56	0.4	45.63	0.5	43.52	0.5	38.69	0.4	42.70	0.5
49.50	0.6	46.28	0.5	46.39	0.5	42.48	0.5	53.17	0.6	52.24	0.6	57.28	0.7	47.50	0.5	47.25	0.5	46.72	0.5
56.13	0.7	55.03	0.6	54.96	0.6	47.54	0.5	56.09	0.7	54.48	0.6	55.65	0.7	45.77	0.5	42.62	0.5	48.20	0.6
64.13	0.8	62.51	0.7	60.24	0.7	55.71	0.7	61.16	0.7	56.76	0.7	57.11	0.7	79.69	1.0	74.18	0.9		
70.56	0.9	69.19	0.8	71.04	0.9	67.933	0.8	107.33	1.4	103.61	1.3	116.07	1.5	119.39	1.5	100.67	1.3		
78.64	1.0	76.00	0.9	76.23	0.9			93.01	1.2	88.61	1.1	86.16	1.1						
86.46	1.1	81.55	1.0	80.59	1.0	76.13	0.9	151.82	2.0	143.00	1.8	143.95	1.9	117.47	1.5	139.42	1.8		
93.10	1.2	87.33	1.1	88.39	1.1	90.49	1.1	109.27	1.4	99.72	1.3	99.00	1.2	120.18	1.5	144.74	1.9	130.02	1.7
100.23	1.3	111.24	1.4	114.11	1.4			107.88	1.4	103.01	1.3	105.61	1.3	112.02	1.4	116.65	1.5	110.48	1.4
108.91	1.4	107.82	1.4	105.60	1.3			112.00	1.4	113.18	1.4	163.58	2.1	119.88	1.5	118.20	1.5	162.40	2.1
113.46	1.4	113.68	1.4	112.19	1.4	122.42	1.6	115.48	1.5	119.64	1.5	174.61	2.3	109.19	1.4	103.69	1.3	124.42	1.6
123.61	1.6	116.71	1.5	117.07	1.5	124.69	1.6	130.68	1.7	129.31	1.7	126.34	1.6	120.32	1.5	129.61	1.7	134.57	1.7
131.07	1.7	118.85	1.5	122.58	1.6	136.31	1.8	146.25	1.9	143.51	1.9	146.60	1.9	147.67	1.9	141.16	1.8	116.69	1.5
138.61	1.8	135.63	1.7	128.86	1.7	142.13	1.8	157.56	2.0	150.41	1.9			153.96	2.0	151.84	2.0	144.51	1.9
146.72	1.9	141.19	1.8	140.82	1.8	147.51	1.9							167.57	2.2	162.25	2.1	140.44	1.8
154.41	2.0	159.57	2.1	157.35	2.0	165.76	2.2												

Table 45 R_2^* values from the reference and from the phantom, measured using the Philips optimised sequence with Compressed SENSE and SENSE, and with the Siemens optimised sequence. R_2^* measurements are given for the large, medium, and small spheres. The R_2^* values have been converted into iron concentration using the Alam[60] and Hankins'[58] equations. The key is:

- same value as reference value
- +/-0.1 mg/g dw from reference value
- +/-0.2 mg/g dw from reference value
- more than +/-0.2 mg/g dw from reference value

Spearman's rho and ICCs show that there was excellent correlation and agreement between the measured values and the reference values for Philips, with similar results for both Compressed SENSE and SENSE. Spearman's rho and ICC were lower for the medium and small spheres compared to the large spheres, but still high. For Siemens, there was also excellent correlation and agreement for the larger spheres, but weaker correlation and agreement for the medium and small spheres. There was small bias which was less than 7.33 s^{-1} (corresponding to 0.1 mg/g dw) for the large spheres for both Philips and Siemens, although the LoA were wider than this. For the medium and small spheres, the biases were larger than 7.33 s^{-1}, and the LoA were wide. The LoA were larger for Siemens compared to Philips.

The results in Table 45 show that, for the large spheres, the equivalent iron concentrations of the measured values were generally either equal to or +/-0.1 mg/g dw of the reference values for Philips, although for Siemens there were three concentrations where the difference was 0.2 mg/g dw from the reference values. For both Philips and Siemens there were multiple examples of the equivalent iron concentrations being more than 0.2 mg/g dw from the reference values in the medium and small spheres. When comparing across Philips Compressed SENSE and SENSE, in terms of differences in iron compared to the reference values, the results were different for different spheres. However, overall, there were similar numbers of measurements which were within each category of difference from the reference value.

Figure 62 (Philips) and Figure 63 (Siemens) show box plots comparing the R_2* measured using the different sequences for concentrations 3, 4 and 5 with the background values. The concentrations correspond to iron of 0.6 mg/g dw, 0.7 mg/g dw and 0.8 mg/g dw, and the background value corresponds to 0.9 mg/g dw. The equivalent iron differences between the concentrations and the background are 0.3 mg/g dw, 0.2 mg/g

dw and 0.1 mg/g dw which are applicable to detecting iron sparing in HCC. Figure 62 and Figure 63 show that for Concentrations 3, 4 and 5 there was no overlap in the interquartile ranges of the ROIs for the spheres and the background R_2*/equivalent iron, for both Philips and Siemens acquisitions. For Concentration 4 the range of R_2* in the sphere did overlap with the R_2* of the background for Philips.

Table 46 shows the differences in the equivalent iron for the spheres with concentrations 1 to 5 compared to the background. Generally, the iron differences between the spheres and the background were the same or 0.1 mg/g dw different to the differences between the reference values and the reference background values, for both Philips and Siemens large spheres, and for Philips medium spheres results. For Siemens, two medium spheres were different by 0.2 mg/g dw to the reference difference. In the small spheres, for Philips there were several spheres with 0.2 mg/g dw difference to the reference difference, and for Siemens there was one sphere with more than 0.2 mg/g dw difference.

Although not directly the purpose of this work, I also compared the results between the Philips and Siemens scanners. These results are given in Appendix D.

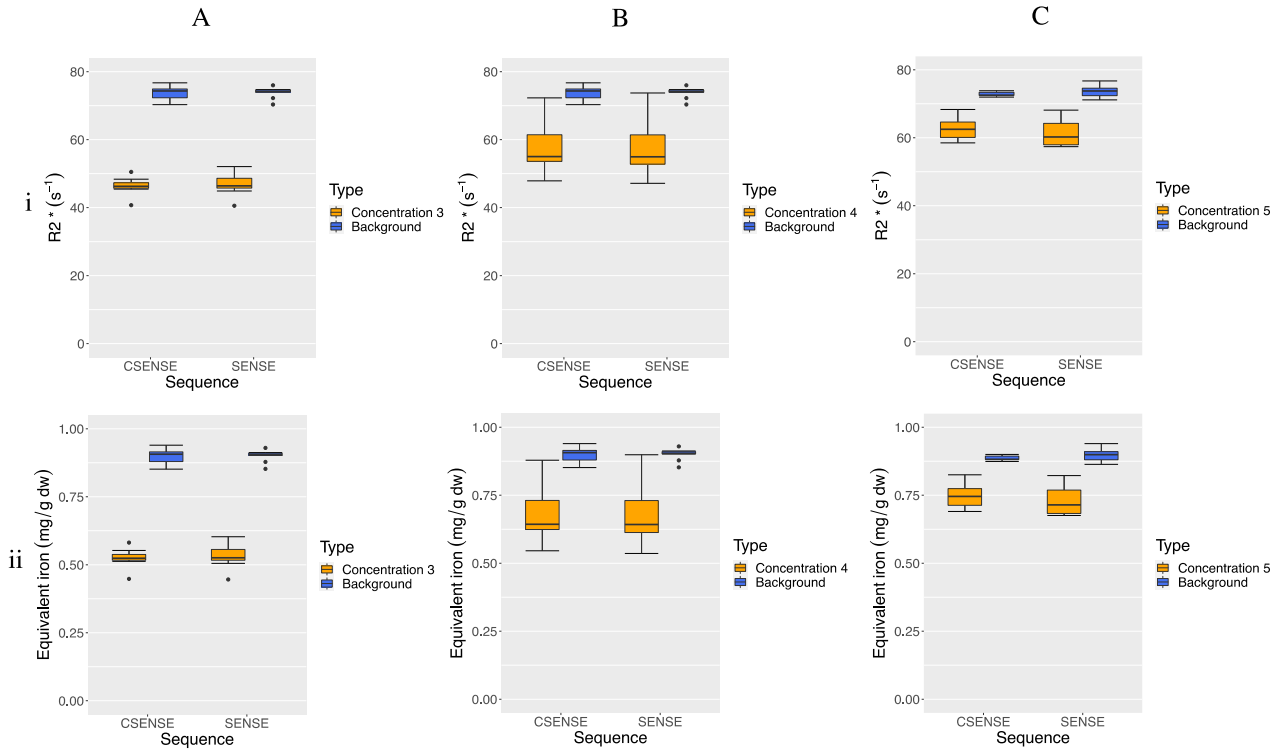


Figure 62 For Philips: Box plots for the (i) R_2* pixels and (ii) Equivalent iron values for the ROIs placed on spheres of different concentrations and the ROIs placed on the background. (A) shows the results for Concentration 3 and background for the upper plate (as that was where Concentration 3 sphere was located, (B) shows the results for Concentration 4 and the background measured for the upper plate (as that was where Concentration 4 sphere was located) and (C) shows the results for Concentration 5 and the background measured for the central plate (as that was where Concentration 5 sphere was located). Results are shown for the large spheres only

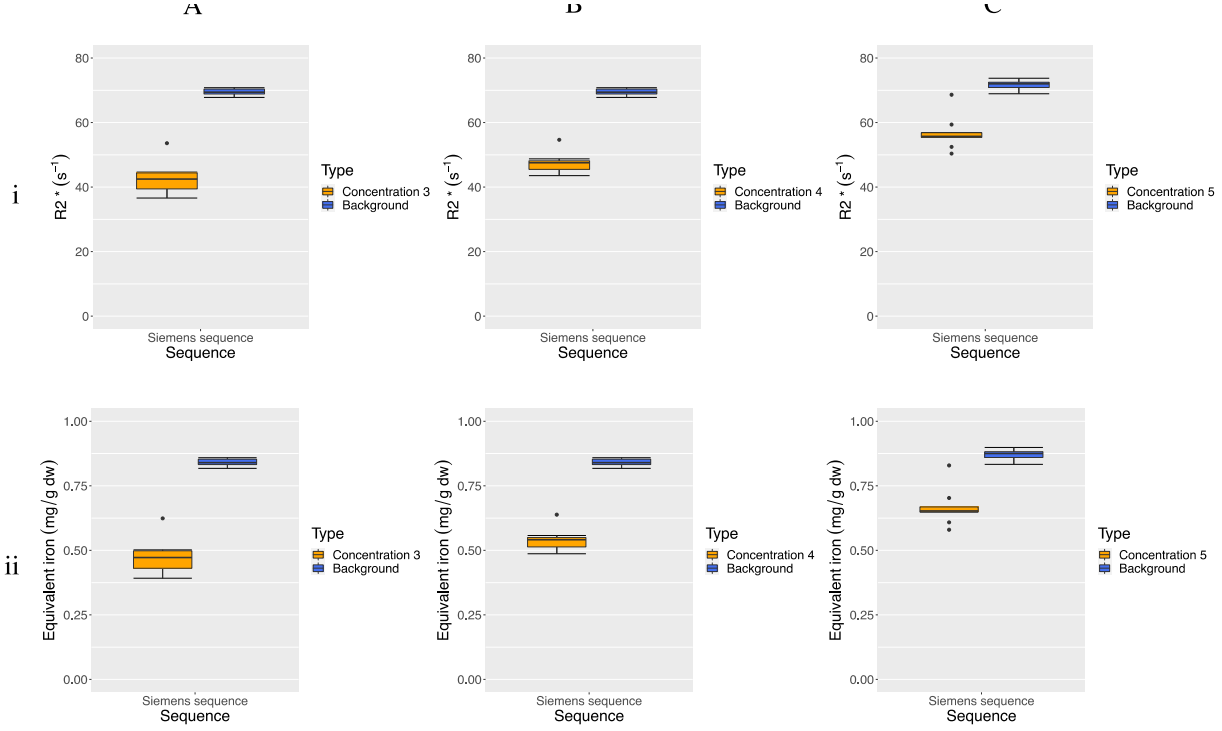
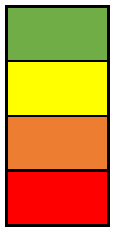


Figure 63 For Siemens: Box plots for the (i) R_2* pixels and (ii) Equivalent iron values for the ROIs placed on spheres of different concentrations and the ROIs placed on the background. (A) shows the results for Concentration 3 and background for the upper plate (as that was where Concentration 3 sphere was located, (B) shows the results for Concentration 4 and the background measured for the upper plate (as that was where Concentration 4 sphere was located) and (C) shows the results for Concentration 5 and the background measured for the central plate (as that was where Concentration 5 sphere was located). Results are shown for the large spheres only

		Large Spheres			Medium Spheres			Small spheres		
Concentration	Reference equivalent iron concentration difference from background (mg/g dw)	Philips Compressed SENSE equivalent iron concentration difference from background (mg/g dw)	Philips SENSE equivalent iron concentration difference from background (mg/g dw)	Siemens equivalent iron concentration difference from background (mg/g dw)	Philips Compressed SENSE equivalent iron concentration difference from background (mg/g dw)	Philips SENSE equivalent iron concentration difference from background (mg/g dw)	Siemens equivalent iron concentration difference from background (mg/g dw)	Philips Compressed SENSE equivalent iron concentration difference from background (mg/g dw)	Philips SENSE equivalent iron concentration difference from background (mg/g dw)	Siemens equivalent iron concentration difference from background (mg/g dw)
1	-0.5	-0.5	-0.5	-0.5	-0.5	-0.5	-0.3	-0.4	-0.4	0.9
2	-0.4	-0.4	-0.4	-0.4	-0.4	-0.5	-0.3	-0.4	-0.5	-0.3
3	-0.3	-0.4	-0.4	-0.3	-0.3	-0.3	-0.1	-0.4	-0.4	-0.3
4	-0.2	-0.3	-0.3	-0.3	-0.2	-0.3	-0.1	-0.4	-0.4	-0.2
5	-0.1	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	+0.1	0.0	N/A

Table 46 Differences in equivalent iron concentration of spheres vs the background values, for spheres with concentrations 1-5. Comparison is made with the difference in reference values for the concentrations vs the background using colour coding:



same difference as reference value difference

+/-0.1 mg/g dw different from reference value difference

+/-0.2 mg/g dw different from reference value difference

more than +/-0.2 mg/g dw f different from reference value difference

9.2.4. Discussion

Creating a phantom to mimic the R_2* of small HCCs and surrounding liver allowed evaluation of several of the requirements which previously had been difficult to assess, in particular detecting small differences in iron concentration in localised areas.

The similarity of the results between the reference values and the measured values of the large spheres, for both Philips sequences and the Siemens sequence (as shown by Spearman's rho, ICC, and the Bland Altman analysis) indicate that we can detect typical R_2* values associated with the amount of iron sparing expected in HCC. Additionally, the differences from the reference values were typically equivalent to 0.1 mg/g dw or less, showing that there is sufficient accuracy to detect small changes in iron, for example if monitoring a nodule over time.

The results also indicate that local differences in iron concentration of 0.2 mg/g dw or more can be detected, given that there was a clear difference in R_2* between the large spheres and background for differences in R_2* equivalent to 0.1 mg/g dw, 0.2 mg/g dw and 0.3 mg/g dw. This is important in iron sparing, as it is the difference in iron between areas in the liver that is important.

The third requirement under consideration was that of detecting tumours of 1 cm or more. The spheres 2 cm in diameter gave the best results, and the smallest spheres (1 cm in diameter) gave the worst results. Early HCC is defined as <3 cm[9] so it is promising that this method could potentially be used to detect iron sparing in a sphere with diameter of 2 cm. However, ideally, very early HCC would be detected i.e. <2 cm[9], as outcomes are best when HCC is detected very early[9]. Given that the results were poor for the 1 cm diameter spheres, it may be necessary to perform further work to optimise the sequences for detecting very small regional changes in R_2*. However, it should be noted that due to the thickness of the plastic the internal diameter of the 1 cm sphere was less than 1 cm

(the external diameter was 1 cm), so it is not surprising that the sequence did not perform so well for the 1 cm spheres, given that we had optimised the sequences specifically for detecting tumours of 1cm or more. Additionally, the boundaries of the spheres do not correctly mimic HCC, and the potential for bubbles and other causes of artifacts make it more challenging to detect R_2* differences than might be the case in HCC. In the SELINA study, the inclusion criterium is "early HCC" (i.e., <3 cm), not "very early HCC" (i.e., <2cm) so most patients will have tumours of 2-3 cm. Therefore, the sequence may be adequate for detecting iron sparing in the tumours of the patients in the SELINA study.

The Siemens scanner gave slightly poorer results in terms of correlation and agreement with the reference values. This may be due to issues which arose as a result of transporting the phantom: "holes" where spheres had previously been and more bubbles being present. Ideally a more robust phantom would be made to compare results from the two scanners.

On the Philips scanner, SENSE and Compressed SENSE gave similar results when compared to the reference values. The results from this study highlight that in a phantom mimicking HCC the results are similar in terms of meeting requirements for detecting iron sparing, and so it may be acceptable to use Compressed sensing in this application.

Given that the Philips and Siemens R_2* results showed strong correlation and agreement and small biases, this indicates that both sequences could potentially be used in the same clinical trial, such as SELINA, and also in situations such as monitoring nodules over time.

Phantoms are frequently used in MRI for assessing sequences, and $MnCl_2$ has been used to create solutions of different R_2* in various commercial phantoms, such as Calimetrix phantoms (one of which was used in earlier parts of this thesis). However, these often comprise of vials and not small spheres. Small spheres are used in phantoms

such as the NIST phantom[108], but the purpose of those phantoms is not to mimic HCC concentrations. By using small spheres to mimic HCCs we could assess spatial variation in R_2*. Additionally, we had reference values obtained from the scanning of larger tubes, which could be used as a ground truth. The results enabled assessment of the requirements for iron sparing which would be difficult to assess in healthy volunteers. However, the phantom did have some limitations, mainly due to practical considerations during the manufacturing process, as outlined in Section 9.2.2, such as the use of $MnCl_2$ to modulate R_2* instead of iron, and not including other substances that might be found in tumours such as fat. Additionally, the phantom did not contain the full range of concentrations which the requirements specified, so it may be advisable to create another phantom, so all concentrations are covered.

In conclusion, this work indicated that the optimised Philips and Siemens sequence could be used to detect localised changes in R_2* associated with iron sparing but potentially not for tumours less than 15 mm.

9.3. Patients with Hepatocellular Carcinoma

9.3.1. Introduction

Following the assessment of the optimised sequences in healthy volunteers and the phantom, it was decided that they would be appropriate for establishing if iron sparing does occur in tumours of around 2 cm or more. There are still some areas for improvement, such as through-slice susceptibility, but these could be mitigated through post-processing. The SELINA study, consisting of 80 patients with early HCC (<3cm) scanned on either Siemens or Philips 3 T scanner, provides an ideal population for scanning to assess the presence of iron sparing. Recruitment to the SELINA study began in May 2022, with the first patient scanned at the end June 2022. This was much later than originally anticipated due to delays associated with Covid-19. However, the delay

enabled a more thorough assessment of the sequences prior to scanning than would otherwise have been feasible, which was beneficial. At the time of writing four patients had been recruited and this section contains case studies of these patients, to show the utility of the sequences in establishing whether iron sparing occurs in tumours, and to what extent.

9.3.2. Methods

Patients were recruited to the SELINA study from May 2022 onwards, with the first four patients recruited by September 2022. All these patients were scanned on the Siemens Prisma 3 T scanner between May 2022 and September 2022 (at the time of writing no patients had been scanned on the Philips 3 T Ingenia scanner). Patients all had early HCC confirmed with radiology (MRI/CT) and/or histology. The scanning for SELINA took place before any treatment of the tumour. The protocol they were scanned with included the optimised sequence given in Section 7.3.4. Three slabs of the liver were acquired to give full liver coverage. Other sequences were also included in the protocol, including T_1-weighted, T_2-weighted and diffusion weighted imaging.

As in all other sections, T_2* and R_2* maps were generated using the MAGO method in MATLAB. The other sequences, such as T_1-weighted and diffusion weighted imaging, were used to identify the location of the tumour and aid segmentation. The additional images were particularly helpful to distinguish vessels from tumour in some cases. I did not have access to the clinical contrast enhanced images and in some cases identifying the tumour was challenging. I manually segmented the tumour and liver (excluding the tumour but including vessels) on the 2D slice where the tumour was most visible and calculated median R_2* of the segmentations. This was converted to iron using Alam's[60] and Hankins'[58] equations, as in previous sections.

To view the volumes in three dimensions, instead of by slice, 3D volumes of the T_2* and R_2* maps were generated in NIFTI format. To make it easier for viewing, the three slabs were combined into one volume. To do this, a Python script was created. As there was an overlap between slabs, the lower slab was used for the lower half of the overlap and the upper slab for the upper half. Examples of the combined volumes are given. In the next section.

9.3.3. Results

Figure 64 shows R_2* maps with segmentation of the tumour and liver, with corresponding median R_2* and iron concentration of the tumour and liver (excluded tumour). Figure 65 and Figure 66 give examples of the combined slabs into one volume.

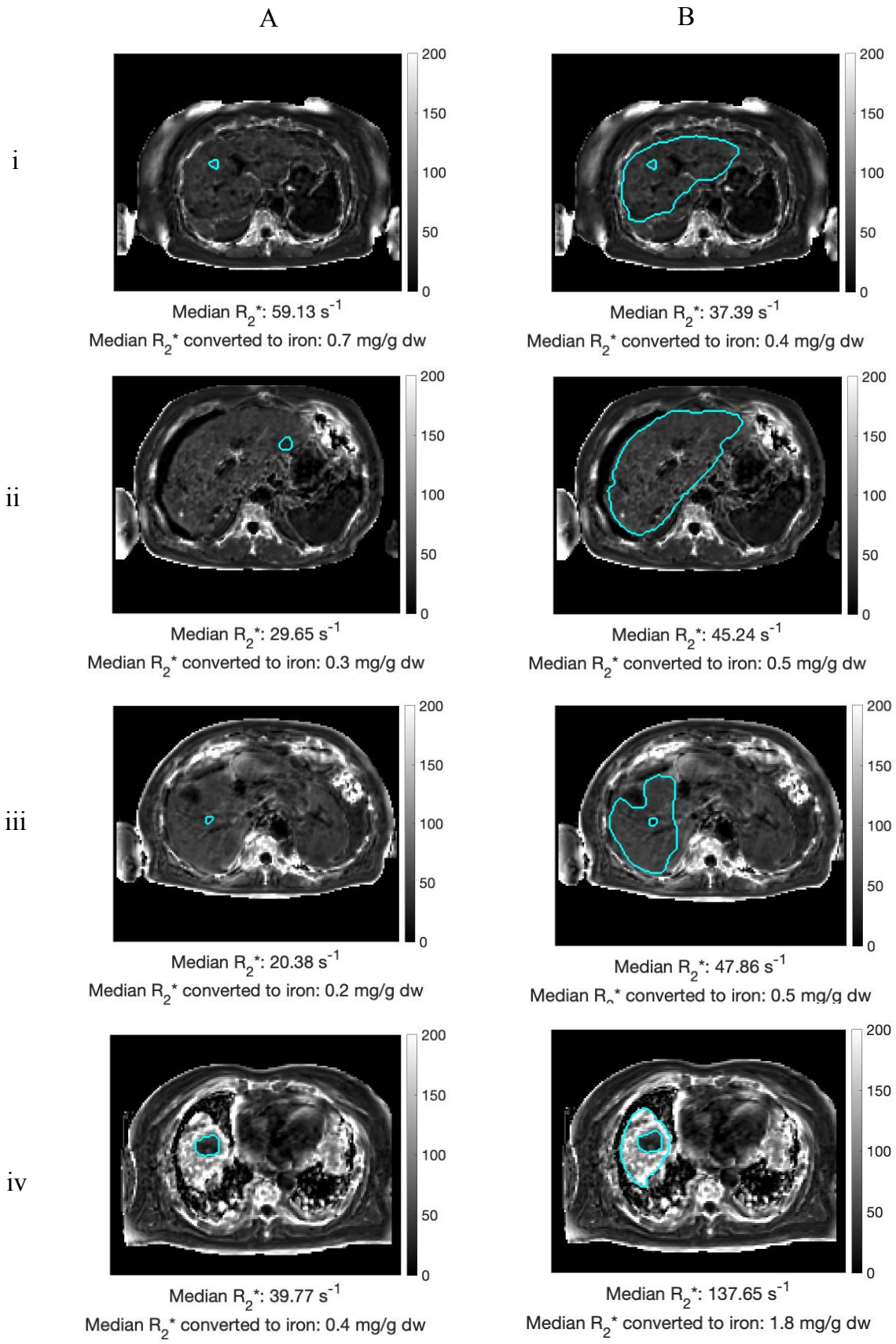


Figure 64 R_2* map with tumour segmentation (A) and R_2* map with liver segmentation (excluding vessels) (B) for Patients (i) 1, (ii) 2, (iii) 3 and (iv) 4. Median R2* of the segmentation, and equivalent iron concentration is given below each map.

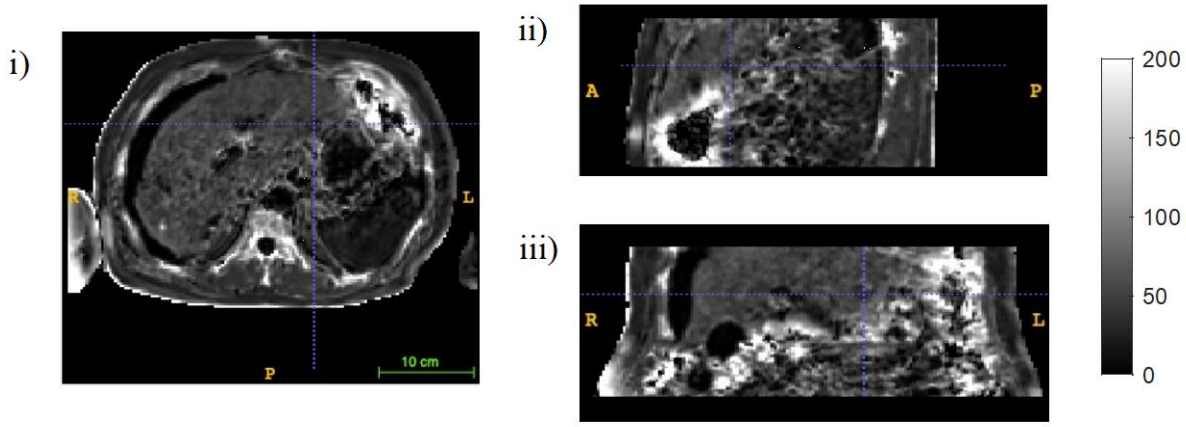


Figure 65 Combined slabs into 3D R_2* map, shown in i) axial, (ii) sagittal and (iii) coronal views for Patient 2.

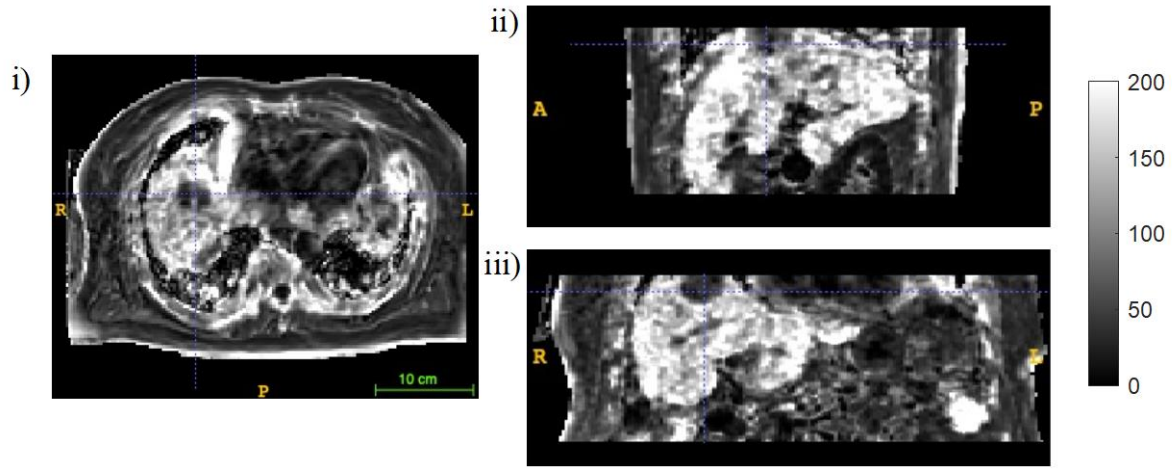


Figure 66 Combined slabs into 3D R_2* map, shown in i) axial, (ii) sagittal and (iii) coronal views for Patient 4.

9.3.4. Discussion

In all four cases, there was a difference in R_2* between the liver and lesion identified as a potential tumour, indicating that R_2* mapping could be useful for early detection of HCC. For three of the four cases the tumour had lower R_2* than the liver, indicating that iron sparing may indeed occur in tumours. These three cases also indicate that the sequence has met the requirements for detecting iron sparing in early tumours as they show that small differences between tumour and liver can be identified (equivalent to 0.2

mg/g dw iron concentration in one case) in small tumours (<3cm). It should be noted, as previously highlighted in Section 3.8.3, that the conversion of R_2* to iron concentration in the tumour using equations such as Hankins'[58] may not be applicable in tumours, however assessing this was outside the scope of this work to date and using the conversion provides a useful indication of the amount of iron, even if not entirely accurate.

For one case, R_2* was higher in the tumour than the liver, which could indicate that the difference in R_2* is reflecting something other than iron concentration, for example hypoxia. As mentioned in Section 3.8.3 deoxyhaemoglobin increases R_2*, so if the tumour hypoxic this could cause an increase in R_2*, even if iron (in the form of ferritin and hemosiderin) decreases[69]. However, the recent narrative review[73] referenced earlier in this thesis stated that so far the results are inconclusive as to whether there is severe hypoxia in HCCs. Therefore, we cannot be sure that is the cause. Another potential explanation is that the tumour does have more iron than the liver tissue. This is feasible, given that a previous study which included histological analysis of iron found that some tumours do have more iron than the liver tissue, although most have less iron[18]. The amount of iron could provide diagnostic or prognostic information and investigating this could be a future area of work.

Analysis of patient data confirmed that using additional images, such as diffusion weighted images, is important to aid with identification of the tumour, particularly when distinguishing tumours from vessels. It therefore confirmed that it is likely that several sequences will be required to enable early detection of HCC, and this will be the focus of the SELINA study.

Combining the slabs into one 3D volume worked well, with no indication that image registration is required. This can therefore be used throughout the SELINA work,

although if there is motion between the acquisition of each slab this could mean image registration will be required.

9.4. Summary

In this chapter I evaluated the optimised sequences in phantoms and patients. The results indicate that the sequences may be suitable for detecting iron sparing, especially in tumours of 2 cm or more. The recruitment of further patients to the SELINA study will allow for a more thorough evaluation of the sequence and of whether iron sparing occurs in HCC.

10. Summary and Outlook

Incidence rates of HCC are increasing in many European countries and the United States[7], and death rates are also increasing[8]. To increase survival, earlier diagnosis is essential. QMRI may enable detection of early changes in the liver associated with HCC. Iron sparing is one such change that MRI can in principle detect. However no optimised sequences were available for this application.

In this thesis I outlined the process of protocol optimisation and then assessed the optimised sequences using data from healthy volunteers, a phantom I developed, and an initial cohort of patients. The sequences developed on the 3 T Siemens and 3 T Philips scanners will now be used for the remainder of the SELINA study, where 80 patients with early HCC will be scanned (40 on each scanner), with the aim of developing tools for early detection of liver cancer. The 1.5 T Siemens sequence may be used in the PEARL study, which will involve the scanning of patients at-risk of HCC, who will be followed prospectively through to possible HCC development, again with the aim to detect early features of HCC. However, although the sequences will be used in this large clinical trial, some further development work may be useful in this area.

10.1. Improvements to R_2* maps

Evaluating the optimised sequences in Chapters 8 and 9 highlighted that there are still some quality issues present. Some, such as patient motion during the scan due to difficulties in breath-holding, can be mitigated by giving clearer training and instructions prior to the scan. Additionally, we have added a shorter breath-hold sequence with fewer slices to the SELINA protocol, for patients who struggle with the breath-hold. However, this might not be sufficient. With computers becoming ever faster and continual

improvement in techniques such as machine learning, MRI technology and techniques continue to develop, which may help accelerate imaging even further. For example the use of deep learning may enable better image reconstruction of sparse images[109] but this may not be advisable when looking for small, local changes in the liver, as if only very limited data is sampled it is more likely that these changes will be missed.

Compressed sensing can significantly speed up the acquisition of MRI scans and thus is useful for reducing the breath-hold length in 3D ME GRE sequences for R_2* mapping. However, acquisitions with and without compressed sensing can give different results in terms of R_2* (see Section 5.3) and also seem to create more artifacts (see Chapter 6 and Section 8.3). Further studies are needed to compare acquisitions with and without compressed sensing. To this end the SELINA study will include sequences with and without compressed sensing.

As well being affected by motion, the quality of the images remained impaired in some participants, near the lungs and bowel, due to susceptibility. One potential method to reduce this would be to reduce the slice thickness further, but this would then make the breath-hold longer. An alternative would be to correct for B_0 inhomogeneity. A B_0 field map can be estimated whilst estimating R_2*[64]. Hernando *et al.* showed that by using the B_0 field map they could correct for B_0 inhomogeneities in the liver in data acquired using 3D ME GRE, and that by doing that the R_2* of slices towards the lungs became much more consistent with the R_2* of slices near the centre of the liver[94]. However, they did note that there was incomplete correction of R_2* near the liver boundary and some slices near the liver dome[94]. This may have been due to in-plane B_0 inhomogeneity, which was not corrected for. Therefore, it may be advisable to add a B_0 correction step to the processing of our 3D R_2* maps and extend it to include in-plane as well as through-plane B_0 correction.

10.2. Phantom development

The phantom I made was very useful, not just for assessing whether the optimised protocol could estimate liver iron sparing with sufficient accuracy, but also for cross-vendor comparisons. However, it had limitations. Firstly, it was easily damaged during transit, with some spheres coming loose. To mitigate this, stronger adhesive is advisable. It also did not truly mimic the liver and tumours, as it used $MnCl_2$ to modulate R_2*. It may be preferable to use iron microspheres and also include oil to mimic the presence of fat, as fat can be present in early HCC[110]. Using this combination has been shown to give results which are representative of liver R_2*[86]. It would be difficult to use iron microspheres and oil in spheres, but further effort could be made to do this. Tumours do contain many other components but using iron microspheres and oil would be the next step in the process to provide a better representation. Thinner plastic may be desirable to reduce the layer between the spheres and the surrounding flood solution, which may better represent the tumour boundary.

10.3. Conversion to iron in tumours

A limitation of this work is that no there is no function proposed in the literature to map R_2* to iron concentration in HCC tumours. Given the heterogeneity in HCCs[111] it is likely that no calibration curve would incorporate all tumours. However, this needs to be confirmed in a biopsy paired study. It may be possible to define a family of calibration curves, indexed by other information available about the tumours e.g., from blood markers. A study in which patients who are having tumour biopsy/resection are imaged with our protocol prior to biopsy, and then iron concentration is quantified using the biopsy, would be very useful.

Even within the parenchyma itself, the conversion of R_2* to iron concentration may not apply in HCC patients, who often have cirrhosis and have different aetiologies to those typically used in calibration studies. For example, the majority of the population in both Wood's and Hankins' studies had thalassemia major or sickle cell disease[57,58]. A biopsy paired study, like the one proposed above, would be useful to establish whether a different calibration curve is required for these patients.

10.4. Alternative to R_2* mapping for detecting iron sparing

Although using R_2* is preferable to using R_2 due to shorter acquisition times, it is still not ideal for iron quantification as multiple factors can influence R_2* aside from susceptibility. In complex environments such as tumours the influence of other factors may have confounding effects on using R_2* to measure iron concentration. Quantitative susceptibility mapping (QSM) gives a more direct way of measuring susceptibility[56]. The same acquisition can be used as the one we have developed (a 3D ME GRE sequence), from which a field map can be estimated. The field map has a known relationship with the underlying magnetic susceptibility distribution and so in theory that susceptibility distribution can be mapped. This is a challenging mathematical problem[53], but developments have been made in this area[112] and it is now possible to create QSM maps of the liver which have good correlation with R_2*[113]. This is still a novel area but given the potential usefulness in detecting iron sparing, it would be interesting to explore QSM results in the SELINA study. To this end we have ensured that we also collect phase images for our R_2* mapping sequences, which are essential for QSM.

10.5. Combining with other MRI parameters

In this work I have primarily focused on creating a sequence for detecting iron sparing. It is unlikely that this on its own will enable early detection of HCC. However, combined

with other parameters it may be a useful marker. One such parameter is fat. During the early stages of HCC development fat may accumulate in the tumour, before regressing as the tumour becomes more progressed[37,110]. Given that proton density fat fraction (PDFF) can be estimated using our sequence, we have also potentially developed a sequence which can give us information about the fat content of the tumour. We purposefully kept the flip angle low, as at higher flip angles fat can be influence by T_1 bias[63]. Therefore, in a few breath-holds we can quantify two parameters which may provide markers of HCC development.

Additionally, in the SELINA protocol we have included numerous other sequences including diffusion and magnetic resonance elastography (MRE). HCCs can have restricted diffusion[114], and have been shown to have significantly higher mean shear stiffness compared to benign lesions and liver tissue[115]. A combination of these sequences may enable early detection. We have endeavoured to have as high a resolution as possible for all sequences and to give as much liver coverage as possible, so as not to miss small tumours located anywhere in the liver.

10.6. Conclusion

This work has provided a foundation for future work in developing sequences which may aid with early detection of HCC, in particular in detecting iron sparing. It will be exciting to see the results of the SELINA study as more patients are recruited.

11. References

1. Columbia Surgery. Liver Functions, Location, Anatomy and Disease | Columbia Surgery. https://columbiasurgery.org/liver/liver-and-its-functions. Published 2022. Accessed August 15, 2022.
2. Abdel-Misih SRZ, Bloomston M. Liver Anatomy. *Surg Clin North Am*. 2010;90(4):643-653. doi:10.1016/j.suc.2010.04.017
3. Anthony PP, Ishak KG, Nayak NC, Poulsen HE, Scheuer PJ, Sobin LH. The morphology of cirrhosis: definition, nomenclature, and classification. *Bull World Health Organ*. 1977;55(4):521-540.
4. McGlynn KA, Petrick JL, London WT. Global Epidemiology of Hepatocellular Carcinoma: An Emphasis on Demographic and Regional Variability. *Clin Liver Dis*. 2015;19(2):223-238. doi:10.1016/j.cld.2015.01.001
5. Sanyal AJ, Yoon SK, Lencioni R. The Etiology of Hepatocellular Carcinoma and Consequences for Treatment. *Oncologist*. 2010;15(S4):14-22. doi:10.1634/theoncologist.2010-s4-14
6. LI-RADS® CT/MRI v2018 | American College of Radiology. https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/LI-RADS/LI-RADS-CT-MRI--v2018. Accessed August 25, 2022.
7. Fitzmaurice C, Akinyemiju T, Abera S, et al. The burden of primary liver cancer and underlying etiologies from 1990 to 2015 at the global, regional, and national level results from the global burden of disease study 2015. *JAMA Oncol*. 2017;3(12):1683-1691. doi:10.1001/jamaoncol.2017.3055
8. Institute NC. Liver and Intrahepatic Bile Duct Cancer - Cancer Stat Facts. *SEER*. 2018. https://seer.cancer.gov/statfacts/html/livibd.html. Accessed August 15, 2022.
9. Galle PR, Forner A, Llovet JM, et al. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J Hepatol*. 2018;69(1):182-236. doi:10.1016/j.jhep.2018.03.019
10. Marrero JA, Kulik LM, Sirlin CB, et al. Diagnosis, Staging, and Management of Hepatocellular Carcinoma: 2018 Practice Guidance by the American Association for the Study of Liver Diseases. *Hepatology*. 2018;68(2):723-750. doi:10.1002/hep.29913
11. Kokudo N, Hasegawa K, Akahane M, et al. Evidence-based Clinical Practice Guidelines for Hepatocellular Carcinoma: The Japan Society of Hepatology 2013 update (3rd JSH-HCC Guidelines). *Hepatol Res*. 2015;45(2):123-127. doi:10.1111/hepr.12464
12. Cholankeril G, Patel R, Khurana S, Satapathy SK. Hepatocellular carcinoma in non-alcoholic steatohepatitis: Current knowledge and implications for management. *World J Hepatol*. 2017;9(11):533. doi:10.4254/WJH.V9.I11.533
13. Tzartzeva K, Obi J, Rich NE, et al. Surveillance Imaging and Alpha Fetoprotein for Early Detection of Hepatocellular Carcinoma in Patients With Cirrhosis: A Meta-analysis. *Gastroenterology*. 2018;154(6):1706. doi:10.1053/J.GASTRO.2018.01.064
14. Schraml C, Kaufmann S, Rempp H, et al. Imaging of HCC-current state of the art. *Diagnostics*. 2015;5(4):513-545. doi:10.3390/diagnostics5040513
15. Tedeschi E, Caranci F, Giordano F, Angelini V, Cocozza S, Brunetti A. Gadolinium retention in the body: what we know and what we can do. *Radiol Medica*. 2017;122(8):589-600. doi:10.1007/s11547-017-0757-3
16. Kim HL, An J, Park JA, Park SH, Lim YS, Lee EK. Magnetic Resonance Imaging Is Cost-Effective for Hepatocellular Carcinoma Surveillance in High-Risk Patients With Cirrhosis. *Hepatology*. 2019;69(4):1599-1613. doi:10.1002/hep.30330
17. Ebara M, Fukuda H, Hatano R, et al. Metal contents in the liver of patients with chronic liver disease caused by hepatitis C virus: Reference to hepatocellular carcinoma. *Oncology*. 2003;65(4):323-330. doi:10.1159/000074645
18. Honda H, Kaneko K, Kanazawa Y, et al. MR imaging of hepatocellular carcinomas: Effect of Cu and Fe contents on signal intensity. *Abdom Imaging*. 1997;22(1):60-66. doi:10.1007/s002619900141
19. Tashiro H, Kawamoto T, Okubo T, Koide O. Variation in the Distribution of Trace Elements in Hepatoma. *Biol Trace Elem Res*. 2003;49.
20. Ebara M, Fukuda H, Kojima Y, et al. Small hepatocellular carcinoma: Relationship of signal intensity

to histopathologic findings and metal content of the tumor and surrounding hepatic parenchyma. *Radiology*. 1999;210(1):81-88. doi:10.1148/radiology.210.1.r99ja4181
21. Honda H, Onitsuka H, Kanazawa Y, et al. MR imaging of hepatocellular carcinoma: Correlation of metal content and signal intensity. *Acta radiol*. 1995;36(2):163-167. doi:10.1177/028418519503600211
22. Henninger B, Alustiza J, Garbowski M, Gandon Y. Practical guide to quantification of hepatic iron with MRI. *Eur Radiol*. 2020;30(1):383-393. doi:10.1007/s00330-019-06380-9
23. Encyclopædia Britannica. https://www.britannica.com/science/liver#/media/1/344579/1087. Accessed August 16, 2022.
24. Encyclopædia Britannica. https://www.britannica.com/science/liver#/media/1/344579/68633. Accessed August 16, 2022.
25. Encyclopædia Britannica. https://www.britannica.com/science/liver#/media/1/344579/60419. Accessed August 16, 2022.
26. Das J. Liver disease pathophysiology. *Clin Pharm*. 2011;3.
27. Rojkind M, Giambrone MA, Biempica L. Collagen Types in Normal and Cirrhotic Liver. *Gastroenterology*. 1979;76(4):710-719. doi:10.1016/S0016-5085(79)80170-5
28. Hytiroglou P, Snover DC, Alves V, et al. Beyond "Cirrhosis." *Am J Clin Pathol*. 2012;137(1):5-9. doi:10.1309/AJCP2T2OHTAPBTMP
29. Pinzani M, Rosselli M, Zuckermann M. Liver cirrhosis. *Best Pract Res Clin Gastroenterol*. 2011;25(2):281-290. doi:10.1016/j.bpg.2011.02.009
30. Garcia-Tsao G, Friedman S, Iredale J, Pinzani M. Now there are many (stages) where before there was one: In search of a pathophysiological classification of cirrhosis. *Hepatology*. 2010;51(4):1445-1449. doi:10.1002/hep.23478
31. Farazi PA, DePinho RA. Hepatocellular carcinoma pathogenesis: From genes to environment. *Nat Rev Cancer*. 2006;6(9):674-687. doi:10.1038/nrc1934
32. Teufel A, Weinmann A, Centner C, et al. Hepatocellular carcinoma in patients with autoimmune hepatitis. *World J Gastroenterol*. 2009;15(5):578. doi:10.3748/wjg.15.578
33. Ascha MS, Hanouneh IA, Lopez R, Tamimi TAR, Feldstein AF, Zein NN. The incidence and risk factors of hepatocellular carcinoma in patients with nonalcoholic steatohepatitis. *Hepatology*. 2010;51(6):1972-1978. doi:10.1002/hep.23527
34. Anstee QM, Reeves HL, Kotsiliti E, Govaere O, Heikenwalder M. From NASH to HCC: current concepts and future challenges. *Nat Rev Gastroenterol Hepatol*. 2019;16(7):411-428. doi:10.1038/s41575-019-0145-7
35. El-Serag HB. Hepatocellular carcinoma. *N Engl J Med*. 2011;365(12):1118-1127. doi:10.1056/NEJMra1001683
36. Di Bisceglie AM. Hepatitis B and hepatocellular carcinoma. *Hepatology*. 2009;49(SUPPL. 5):S56. doi:10.1002/hep.22962
37. Choi JY, Lee JM, Sirlin CB. CT and MR imaging diagnosis and staging of hepatocellular carcinoma: Part I. Development, growth, and spread: Key pathologic and imaging aspects. *Radiology*. 2014;272(3):635-654. doi:10.1148/radiol.14132361
38. Choi JY, Lee JM, Sirlin CB. CT and MR imaging diagnosis and staging of hepatocellular carcinoma. Part II. Extracellular agents, hepatobiliary agents, and ancillary imaging features. *Radiology*. 2014;273(1):30-50. doi:10.1148/radiol.14132362
39. Anderson ER, Shah YM. Iron homeostasis in the liver. *Compr Physiol*. 2013;3(1):315. doi:10.1002/CPHY.C120016
40. Cappellini MD, Farmakis D, Porter J, Taher A. *Guidelines for the Management of Transfusion Dependent Thalassaemia (4th Edition – Version 2.0)*.; 2021. https://thalassaemia.org.cy/wp-content/uploads/2021/06/GUIDELINE-4th-DIGITAL-BY-PAGE.pdf.
41. Paganoni R, Lechel A, Spasic MV. Iron at the interface of hepatocellular carcinoma. *Int J Mol Sci*. 2021;22(8):1-15. doi:10.3390/ijms22084097
42. Sorrentino P, D'Angelo S, Ferbo U, Micheli P, Bracigliano A, Vecchione R. Liver iron excess in patients with hepatocellular carcinoma developed on non-alcoholic steato-hepatitis. *J Hepatol*. 2009;50(2):351-357. doi:10.1016/j.jhep.2008.09.011

43. Pietrangelo A. Iron in NASH, chronic liver diseases and HCC: How much iron is too much? *J Hepatol*. 2009;50(2):249-251. doi:10.1016/j.jhep.2008.11.011
44. Li RK, Palmer SL, Zeng MS, et al. Detection of Endogenous Iron Reduction during Hepatocarcinogenesis at Susceptibility-Weighted MR Imaging: Value for Characterization of Hepatocellular Carcinoma and Dysplastic Nodule in Cirrhotic Liver. *PLoS One*. 2015;10(11). doi:10.1371/journal.pone.0142882
45. Breuer FA, Blaimer M, Mueller MF, et al. Controlled aliasing in volumetric parallel imaging (2D CAIPIRINHA). *Magn Reson Med*. 2006;55(3):549-556. doi:10.1002/mrm.20787
46. Deshmane A, Gulani V, Griswold MA, Seiberlich N. Parallel MR imaging. *J Magn Reson Imaging*. 2012;36(1):55-72. doi:10.1002/jmri.23639
47. Glockner JF, Hu HH, Stanley DW, Angelos L, King K. Parallel MR imaging: A user's guide. *Radiographics*. 2005;25(5):1279-1297. doi:10.1148/rg.255045202
48. Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn Reson Med*. 2007;58(6):1182-1195. doi:10.1002/mrm.21391
49. Geethanath S, Reddy R, Konar AS, et al. Compressed Sensing MRI: A Review. *Crit Rev Biomed Eng*. 2013;41(3):183-204.
50. Delattre BMA, Boudabbous S, Hansen C, Neroladaki A, Hachulla AL, Vargas MI. Compressed sensing MRI of different organs: ready for clinical daily practice? *Eur Radiol*. 2020;30(1):308-319. doi:10.1007/s00330-019-06319-0
51. Sartoretti T, Reischauer C, Sartoretti E, Binkert C, Najafi A, Sartoretti-Schefer S. Common artefacts encountered on images acquired with combined compressed sensing and SENSE. *Insights Imaging*. 2018;9(6):1107-1115. doi:10.1007/s13244-018-0668-4
52. Sharma SD, Fong CL, Tzung BS, Law M, Nayak KS. Clinical image quality assessment of accelerated magnetic resonance neuroimaging using compressed sensing. *Invest Radiol*. 2013;48(9):638-645. doi:10.1097/RLI.0B013E31828A012D
53. Hernando D, Levin YS, Sirlin CB, Reeder SB. Quantification of liver iron with MRI: State of the art and remaining challenges. *J Magn Reson Imaging*. 2014;40(5):1003-1021. doi:10.1002/jmri.24584
54. Henninger B, Alustiza J, Garbowski M, Gandon Y. Practical guide to quantification of hepatic iron with MRI. *Eur Radiol*. 2020;30(1):383-393. doi:10.1007/s00330-019-06380-9
55. St. Pierre TG, Clark PR, Chua-Anusorn W, et al. Noninvasive measurement and imaging of liver iron concentrations using proton magnetic resonance. *Blood*. 2005;105(2):855-861. doi:10.1182/blood-2004-01-0177
56. Labranche R, Gilbert G, Cerny M, et al. Liver iron quantification with MR imaging: A primer for radiologists. *Radiographics*. 2018;38(2):392-412. doi:10.1148/rg.2018170079
57. Wood JC, Enriquez C, Ghugre N, et al. MRI R2 and R2* mapping accurately estimates hepatic iron concentration in transfusion-dependent thalassemia and sickle cell disease patients. *Blood*. 2005;106(4):1460-1465. doi:10.1182/blood-2004-10-3982
58. Hankins JS, McCarville MB, Loeffler RB, et al. R2* magnetic resonance imaging of the liver in patients with iron overload. *Blood*. 2009;113(20):4853-4855. doi:10.1182/blood-2008-12-191643
59. Henninger B, Zoller H, Rauch S, et al. R2∗ relaxometry for the quantification of hepatic iron overload: Biopsy-based calibration and comparison with the literature. *RoFo Fortschritte auf dem Gebiet der Rontgenstrahlen und der Bildgeb Verfahren*. 2015;187(6):472-479. doi:10.1055/s-0034-1399318
60. Alam MH, Auger D, McGill LA, et al. Comparison of 3 T and 1.5 T for T2∗magnetic resonance of tissue iron. *J Cardiovasc Magn Reson*. 2016;18(1). doi:10.1186/s12968-016-0259-9
61. Meloni A, Positano V, Keilberg P, et al. Feasibility, reproducibility, and reliability for the T2* iron evaluation at 3 T in comparison with 1.5 T. *Magn Reson Med*. 2012;68(2):543-551. doi:10.1002/mrm.23236
62. Storey P, Thompson AA, Carqueville CL, Wood JC, De Freitas RA, Rigsby CK. R2* imaging of transfusional iron burden at 3T and comparison with 1.5T. *J Magn Reson Imaging*. 2007;25(3):540-547. doi:10.1002/jmri.20816
63. Reeder SB, Sirlin CB. Quantification of liver fat with magnetic resonance imaging. *Magn Reson Imaging Clin N Am*. 2010;18(3):337-357. doi:10.1016/j.mric.2010.08.013

64. Reeder SB, Pineda AR, Wen Z, et al. Iterative decomposition of water and fat with echo asymmetry and least-squares estimation (IDEAL): Application with fast spin-echo imaging. *Magn Reson Med*. 2005;54(3):636-644. doi:10.1002/mrm.20624
65. Triay Bagur A, Hutton C, Irving B, Gyngell ML, Robson MD, Brady M. Magnitude-intrinsic water–fat ambiguity can be resolved with multipeak fat modeling and a multipoint search method. *Magn Reson Med*. 2019;82(1):460-475. doi:10.1002/mrm.27728
66. Terada T, Kadoya M, Nakanuma Y, Matsui O. Iron -Accumulating Adenomatous Hyperplastic Nodule With Malignant Foci in the Cirrhotic Liver. *Cancer*. 1990;64(9):1994-2000.
67. Hardie AD, Romano PB. The use of T2*-weighted multi-echo GRE imaging as a novel method to diagnose hepatocellular carcinoma compared with gadolinium-enhanced MRI: A feasibility study. *Magn Reson Imaging*. 2010;28(2):281-285. doi:10.1016/j.mri.2009.12.010
68. Hardie AD, Morgan AE. Imaging Requirements for Utilization of T2∗-Weighted Magnetic Resonance Imaging for Identification of Hepatocellular Carcinoma in Cirrhosis: Effect of Hepatic Iron Content. *J Comput Assist Tomogr*. 2015;39(4):468-472. doi:10.1097/RCT.0000000000000252
69. O'Connor JPB, Robinson SP, Waterton JC. Imaging tumour hypoxia with oxygen-enhanced MRI and BOLD MRI. *Br J Radiol*. 2019;92(1096). doi:10.1259/bjr.20180642
70. Karlsson M, Ekstedt M, Dahlström N, et al. Liver R2* is affected by both iron and fat: A dual biopsy-validated study of chronic liver disease. *J Magn Reson Imaging*. 2019;50(1):325-333. doi:10.1002/jmri.26601
71. Sun M, Wang S, Song Q, et al. Utility of R 2 * Obtained from T 2 *-Weighted Imaging in Differentiating Hepatocellular Carcinomas from Cavernous Hemangiomas of the Liver Conclusions: R 2 * is a significant MRI biomarker to differentiate HCC from CHL with satisfying sensitivity and specificity. 2014. doi:10.1371/journal.pone.0091751
72. Shi GZ, Chen H, Zeng WK, et al. R2* value derived from multi-echo Dixon technique can aid discrimination between benign and malignant focal liver lesions. *World J Gastroenterol*. 2021;27(12):1182-1193. doi:10.3748/wjg.v27.i12.1182
73. Cramer T, Vaupel P. Severe hypoxia is a typical characteristic of human hepatocellular carcinoma: Scientific fact or fallacy? *J Hepatol*. 2022;76(4):975-980. doi:10.1016/j.jhep.2021.12.028
74. Pakbaz Z, Fischer R, Vichinsky E, et al. Liver Iron Measurement by SQUID Biosusceptometry Compared to Liver Biopsy: A More Accurate Definition of Optimal Iron Range. *Blood*. 2007;110(11):2675-2675. doi:10.1182/blood.v110.11.2675.2675
75. Garbowski MW, Carpenter JP, Smith G, et al. Biopsy-based calibration of T2* magnetic resonance for estimation of liver iron concentration and comparison with R2 Ferriscan. *J Cardiovasc Magn Reson*. 2014;16(1):1-11. doi:10.1186/1532-429X-16-40
76. d'Assignies G, Paisant A, Bardou-Jacquet E, et al. Non-invasive measurement of liver iron concentration using 3-Tesla magnetic resonance imaging: validation against biopsy. *Eur Radiol*. 2018;28(5):2022-2030. doi:10.1007/s00330-017-5106-3
77. Taher AT, Porter J, Viprakasit V, et al. Deferasirox reduces iron overload significantly in nontransfusion-dependent thalassemia: 1-Year results from a prospective, randomized, double-blind, placebo-controlled study. *Blood*. 2012;120(5):970-977. doi:10.1182/blood-2012-02-412692
78. Barry M, Sherlock S. Measurement of Liver-Iron Concentration in Needle-Biopsy Specimens. *Lancet*. 1971;297(7690):100-103. doi:10.1016/S0140-6736(71)90838-5
79. Kukuk GM, Hittatiya K, Sprinkart AM, et al. Comparison between modified Dixon MRI techniques, MR spectroscopic relaxometry, and different histologic quantification methods in the assessment of hepatic steatosis. *Eur Radiol*. 2015;25(10):2869-2879. doi:10.1007/s00330-015-3703-6
80. Henninger B, Zoller H, Kannengiesser S, Zhong X, Jaschke W, Kremser C. 3D Multiecho Dixon for the Evaluation of Hepatic Iron and Fat in a Clinical Setting. *J Magn Reson Imaging*. 2017;46(3):793-800. doi:10.1002/jmri.25630
81. Henninger B, Plaikner M, Zoller H, et al. Performance of different Dixon-based methods for MR liver iron assessment in comparison to a biopsy-validated R2* relaxometry method. *Eur Radiol*. 2021;31(4):2252-2262. doi:10.1007/s00330-020-07291-w
82. Serai SD, Smith EA, Trout AT, Dillman JR. Agreement between manual relaxometry and semi-automated scanner-based multi-echo Dixon technique for measuring liver T2* in a pediatric and

young adult population. *Pediatr Radiol*. 2018;48(1):94-100. doi:10.1007/s00247-017-3990-y
83. Bulanov PA, Nazarova EE, Menshchikov PE, Kupriyanov DA, Tereshchenko G V., Novichkova GA. Creation of an MRI phantom based on super-paramagnetic nanoparticles of iron oxide (SPIO) for standardization of the conversion of T2 * values in the concentration of iron in the liver. *J Phys Conf Ser*. 2021;2015(1). doi:10.1088/1742-6596/2015/1/012025
84. Mobini N, Malekzadeh M, Haghighatkhah H, Saligheh Rad H. A hybrid (iron–fat–water) phantom for liver iron overload quantification in the presence of contaminating fat using magnetic resonance imaging. *Magn Reson Mater Physics, Biol Med*. 2020;33(3):385-392. doi:10.1007/s10334-019-00795-7
85. Bashir MR, Wolfson T, Gamst AC, et al. Hepatic R2* is more strongly associated with proton density fat fraction than histologic liver iron scores in patients with nonalcoholic fatty liver disease. *J Magn Reson Imaging*. 2019;49(5):1456-1466. doi:10.1002/jmri.26312
86. Zhao R, Hamilton G, Brittain JH, Reeder SB, Hernando D. Design and evaluation of quantitative MRI phantoms to mimic the simultaneous presence of fat, iron, and fibrosis in the liver. *Magn Reson Med*. 2021;85(2):734-747. doi:10.1002/mrm.28452
87. Boyarko AC, Dillman JR, Tkach JA, Pednekar AS, Trout AT. Comparison of compressed SENSE and SENSE for quantitative liver MRI in children and young adults. *Abdom Radiol*. 2021;46(10):4567-4575. doi:10.1007/s00261-021-03092-x
88. Sartoretti T, Reischauer C, Sartoretti E, Binkert C, Najafi A, Sartoretti-Schefer S. Common artefacts encountered on images acquired with combined compressed sensing and SENSE. *Insights Imaging*. 2018;9(6):1107-1115. doi:10.1007/S13244-018-0668-4/TABLES/3
89. Berg RC, Leutritz T, Weiskopf N, Preibisch C. Multi-parameter quantitative mapping of R1, R2*, PD, and MTsat is reproducible when accelerated with Compressed SENSE. *Neuroimage*. 2022;253:119092. doi:10.1016/j.neuroimage.2022.119092
90. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol 9351. Springer, Cham; 2015:234-241. doi:10.1007/978-3-319-24574-4_28
91. Canny J. A Computational Approach to Edge Detection. *IEEE Trans Pattern Anal Mach Intell*. 1986;PAMI-8(6):679-698. doi:10.1109/TPAMI.1986.4767851
92. İdilman İS, Akata D, Özmen MN, Karçaaltıncaba M. Different forms of iron accumulation in the liver on MRI. *Diagnostic Interv Radiol*. 2016;22(1):22-28. doi:10.5152/dir.2015.15094
93. Roberts NT, Hinshaw LA, Colgan TJ, Ii T, Hernando D, Reeder SB. B0 and B1 inhomogeneities in the liver at 1.5 T and 3.0 T. *Magn Reson Med*. 2021;85(4):2212-2220. doi:10.1002/mrm.28549
94. Hernando D, Vigen KK, Shimakawa A, Reeder SB. R2* mapping in the presence of macroscopic B0 field variations. *Magn Reson Med*. 2012;68(3):830-840. doi:10.1002/mrm.23306
95. Obrzut M, Atamaniuk V, Glaser KJ, et al. Value of liver iron concentration in healthy volunteers assessed by MRI. *Sci Reports 2020 101*. 2020;10(1):1-8. doi:10.1038/s41598-020-74968-z
96. Dietrich O, Raya JG, Reeder SB, Reiser MF, Schoenberg SO. Measurement of signal-to-noise ratios in MR images: Influence of multichannel coils, parallel imaging, and reconstruction filters. *J Magn Reson Imaging*. 2007;26(2):375-385. doi:10.1002/JMRI.20969
97. Crameri F, Shephard GE, Heron PJ. The misuse of colour in science communication. *Nat Commun*. 2020;11(1):5444. doi:10.1038/s41467-020-19160-7
98. Crameri F. Scientific colour maps. *Zenodo*. 2020;(2018):1-16. doi:10.5281/ZENODO.5501399
99. Kühn JP, Jahn C, Hernando D, et al. T1 Bias in Chemical Shift-Encoded Liver Fat-Fraction: Role of the Flip Angle. *J Magn Reson Imaging*. 2014;40(4):875. doi:10.1002/JMRI.24457
100. Hernando D, Kramer JH, Reeder SB. Multipeak fat-corrected complex R2* relaxometry: Theory, optimization, and clinical validation. *Magn Reson Med*. 2013;70(5):1319-1331. doi:10.1002/mrm.24593
101. Ludwig J, Hashimoto E, Porayko MK, Moyer TP, Baldus WP. Hemosiderosis in cirrhosis: A study of 447 native livers. *Gastroenterology*. 1997;112(3):882-888. doi:10.1053/gast.1997.v112.pm9041250
102. Hood MN, Ho VB, Smirniotopoulos JG, Szumowski J. Chemical shift: The artifact and clinical tool revisited. *Radiographics*. 1999;19(2):357-371. doi:10.1148/radiographics.19.2.g99mr07357

103. Reeder SB, Cruite I, Hamilton G, Sirlin CB. Quantitative assessment of liver fat with magnetic resonance imaging and spectroscopy. *J Magn Reson Imaging*. 2011;34(4):729-749. doi:10.1002/jmri.22580
104. Westwood MA, Anderson LJ, Firmin DN, et al. Interscanner Reproducibility of Cardiovascular Magnetic Resonance T2* Measurements of Tissue Iron in Thalassemia. *J Magn Reson Imaging*. 2003;18:616-620. doi:10.1002/jmri.10396
105. Kirk P, He T, Anderson LJ, et al. International Reproducibility of Single Breathhold T2* MR for Cardiac and Liver Iron Assessment Among Five Thalassemia Centers. *J Magn Reson Imaging*. 2010;32:315-319. doi:10.1002/jmri.22245
106. Hernando D, Zhao R, Yuan Q, et al. Multi-Center, Multi-Vendor Calibration and Validation of Liver Iron Quantification using R2 * -MRI at 1 .5T and 3T. In: *International Society for Magnetic Resonance in Medicine*. ; 2022.
107. Anstee QM, Reeves HL, Kotsiliti E, Govaere O, Heikenwalder M. From NASH to HCC: current concepts and future challenges. *Nat Rev Gastroenterol Hepatol*. 2019;16(7):411-428. doi:10.1038/s41575-019-0145-7
108. Stupic KF, Ainslie M, Boss MA, et al. A standard system phantom for magnetic resonance imaging. *Magn Reson Med*. 2021;86(3):1194-1211. doi:10.1002/mrm.28779
109. Chandra SS, Bran Lorenzana M, Liu X, Liu S, Bollmann S, Crozier S. Deep learning in magnetic resonance image reconstruction. *J Med Imaging Radiat Oncol*. 2021;65(5):564-577. doi:10.1111/1754-9485.13276
110. Kutami R, Nakashima Y, Nakashima O, Shiota K, Kojiro M. Pathomorphologic study on the mechanism of fatty change in small hepatocellular carcinoma of humans. *J Hepatol*. 2000;33(2):282-289. doi:10.1016/S0168-8278(00)80369-4
111. Schlageter M, Terracciano LM, D'Angelo S, Sorrentino P. Histopathology of hepatocellular carcinoma. *World J Gastroenterol*. 2014;20(43):15955-15964. doi:10.3748/wjg.v20.i43.15955
112. Wang Y, Liu T. Quantitative susceptibility mapping (QSM): Decoding MRI data for a tissue magnetic biomarker. *Magn Reson Med*. 2015;73(1):82-101. doi:10.1002/mrm.25358
113. Sharma SD, Hernando D, Horng DE, Reeder SB. Quantitative susceptibility mapping in the abdomen as an imaging biomarker of hepatic iron overload. *Magn Reson Med*. 2015;74(3):673-683. doi:10.1002/mrm.25448
114. Gluskin JS, Chegai F, Monti S, Squillaci E, Mannelli L. Hepatocellular carcinoma and diffusion-weighted MRI: Detection and evaluation of treatment response. *J Cancer*. 2016;7(11):1565-1570. doi:10.7150/jca.14582
115. Venkatesh SK, Yin M, Glockner JF, et al. MR elastography of liver tumors: Preliminary results. *Am J Roentgenol*. 2008;190(6):1534-1540. doi:10.2214/AJR.07.3123

Appendix A Additional Figures from Chapter 4

Comparisons of 1.5 T-R_2* LIC and 3 T-R_2* LIC Conversion

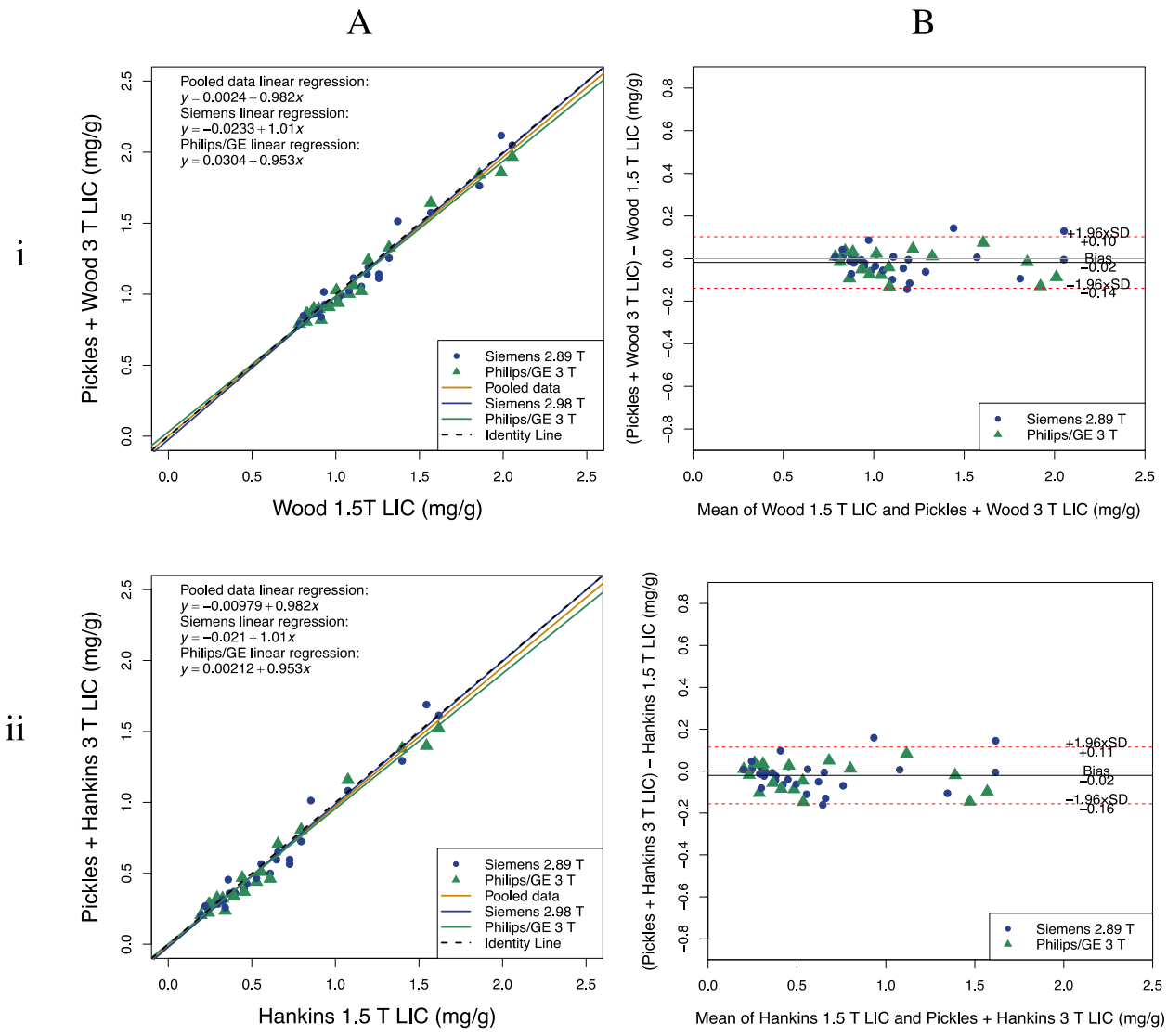


Figure 67 (A) Scatter plot and (B) Bland Altman plot comparing (i) 3 T R_2* converted to LIC via this study R_2* conversion and 1.5 T R_2* converted to LIC via Wood) and (ii)3 T R_2* converted to LIC via this study R_2* conversion and Hankins and 1.5 T R_2* converted to LIC via Hankins. Data point colour depends on the field strength (2.89 T or 3 T), and the three regression lines are for all the data, 2.89 T data and 3 T data.

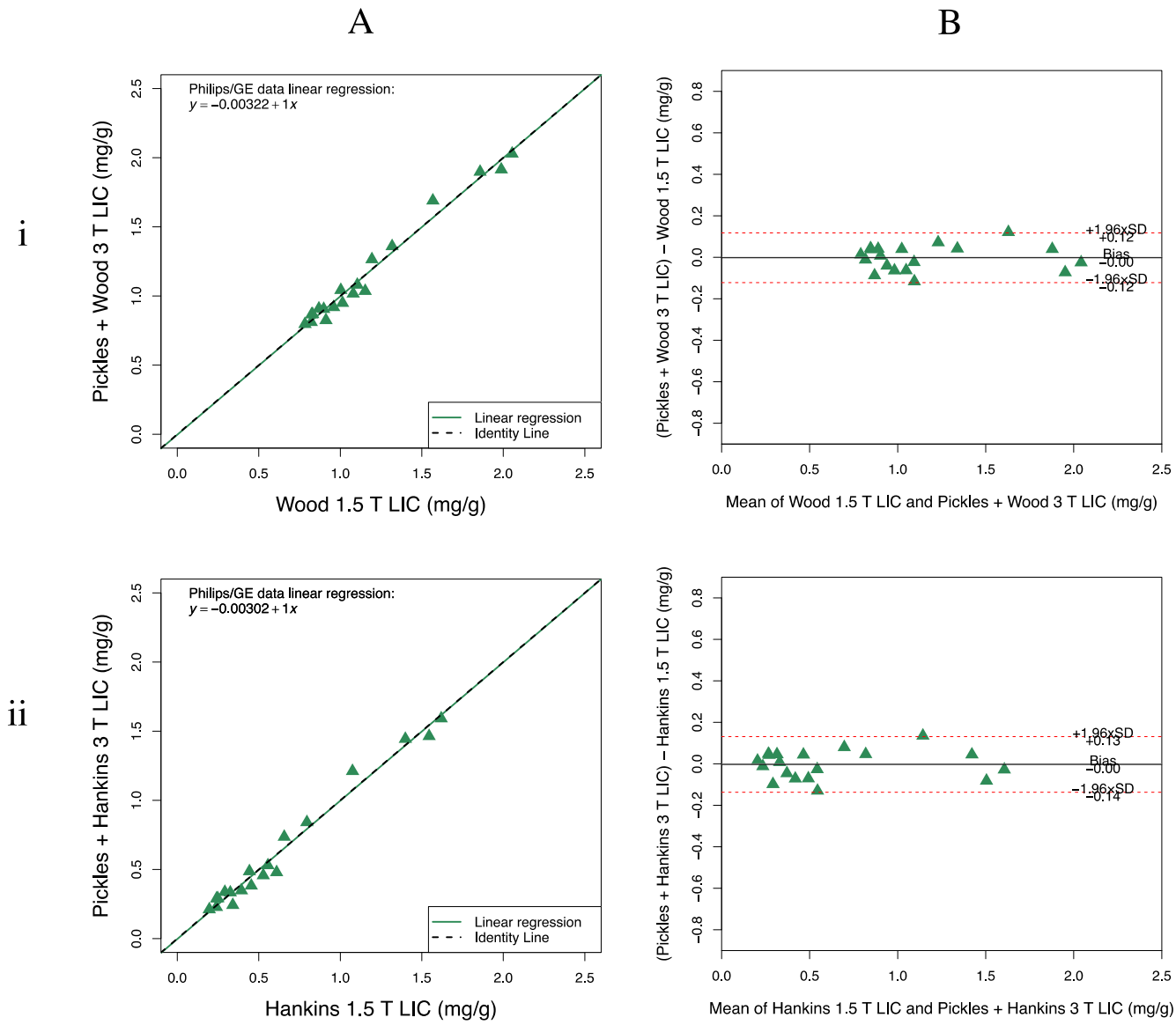


Figure 68 (A) Scatter plot and (B) Bland Altman plot comparing (i) 3 T R_2* converted to LIC via this study's Philips/GE 3 T conversion and Wood and 1.5 T R_2* converted to LIC via Wood and (ii) 3 T R_2* converted to LIC via this study's Philips/GE 3 T conversion and Hankins and 1.5 T R_2* converted to LIC via Hankins. Only Philips/GE data are included in these plots.

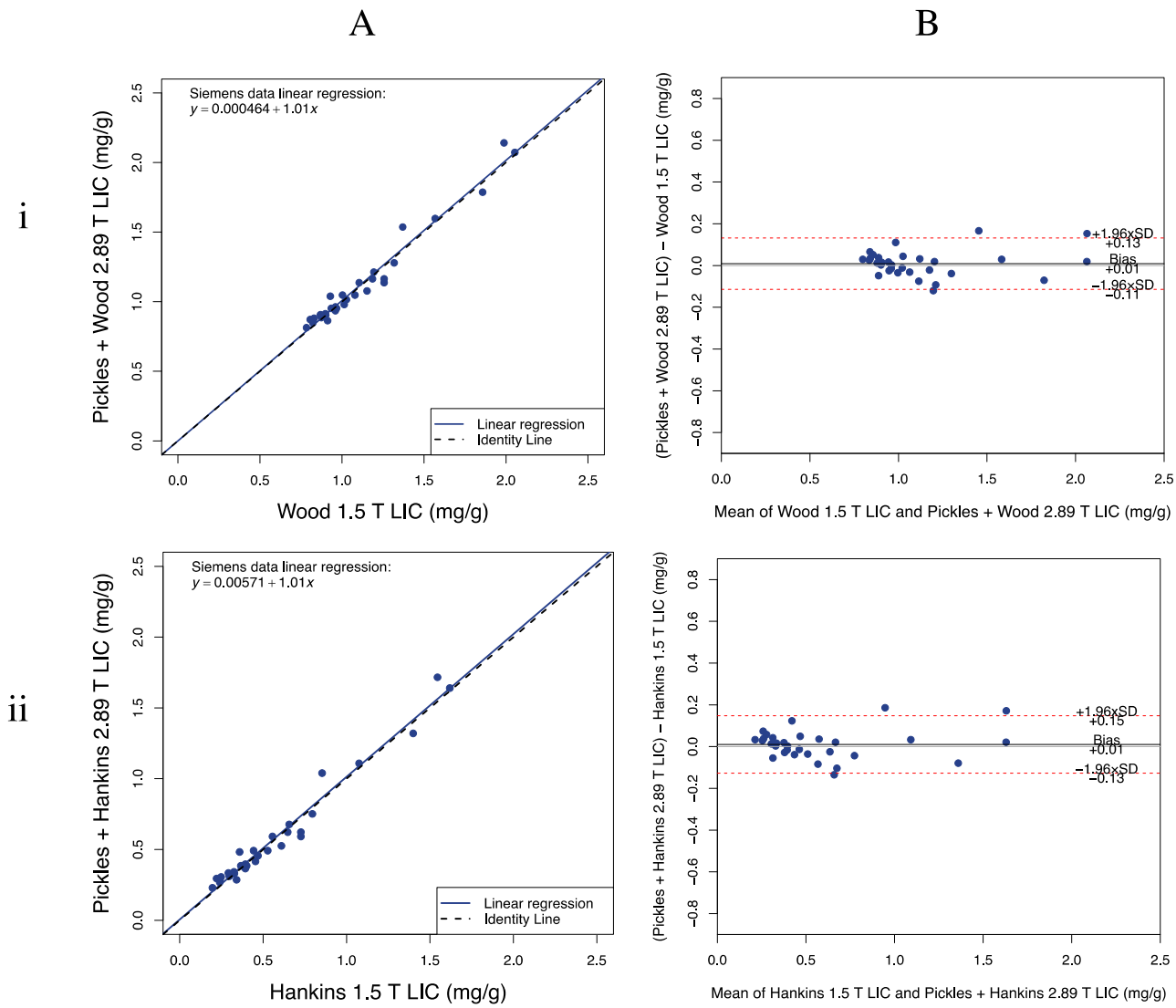


Figure 69 (A) Scatter plot and (B) Bland Altman plot comparing (i) 3 T R_2* converted to LIC via this study's Siemens 2.89 T conversion and Wood and 1.5 T R_2* converted to LIC via Wood and (ii) 3 T R_2* converted to LIC via this study's Siemens 2.89 T conversion and Hankins and 1.5 T R_2* converted to LIC via Hankins. Only Siemens 2.89 T data are included in these plots

Appendix B Additional Figures from Chapter 6

Comparisons of central slice R_2* with other slices in the liver

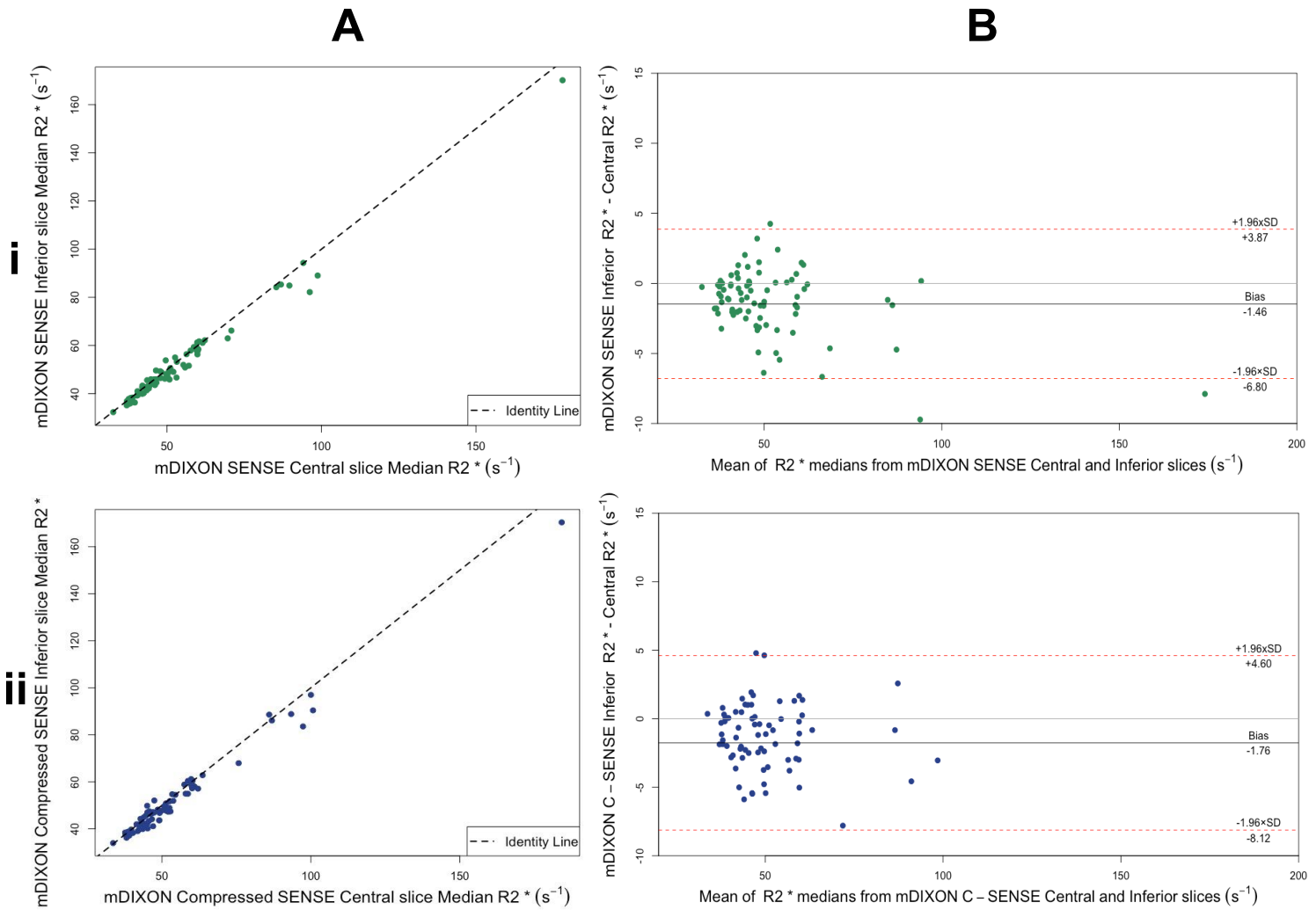


Figure 70 (A) Scatter plot (B) Bland Altman plots, comparing the R2* median of the segmentations of the central slice and inferior slice (i) For mDIXON SENSE and (ii) mDIXON Compressed SENSE

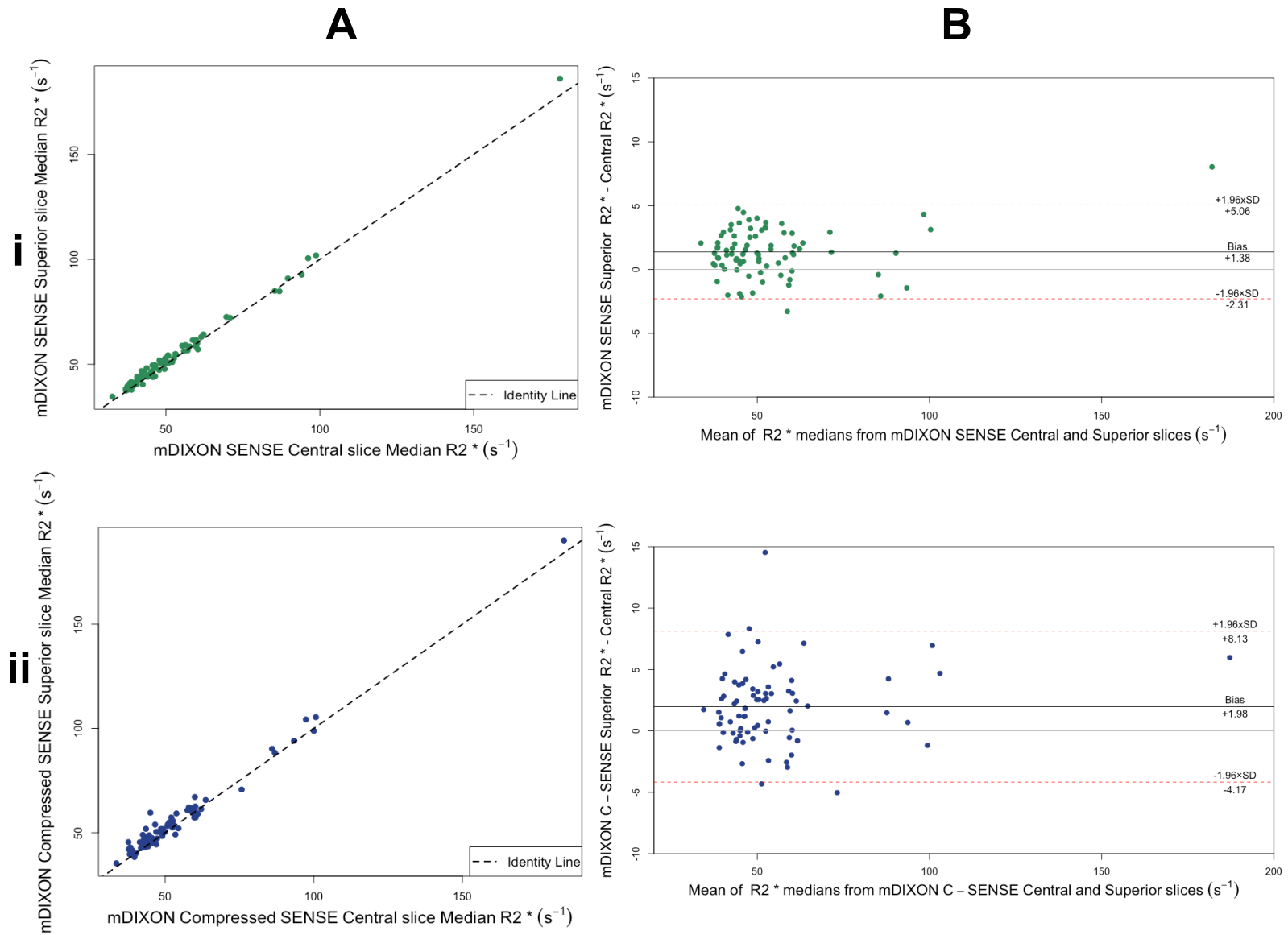


Figure 71 (A) Scatter plot (B) Bland Altman plots, comparing the R_2* median of the segmentations of the central slice and superior slice (i) For mDIXON SENSE and (ii) mDIXON Compressed SENSE

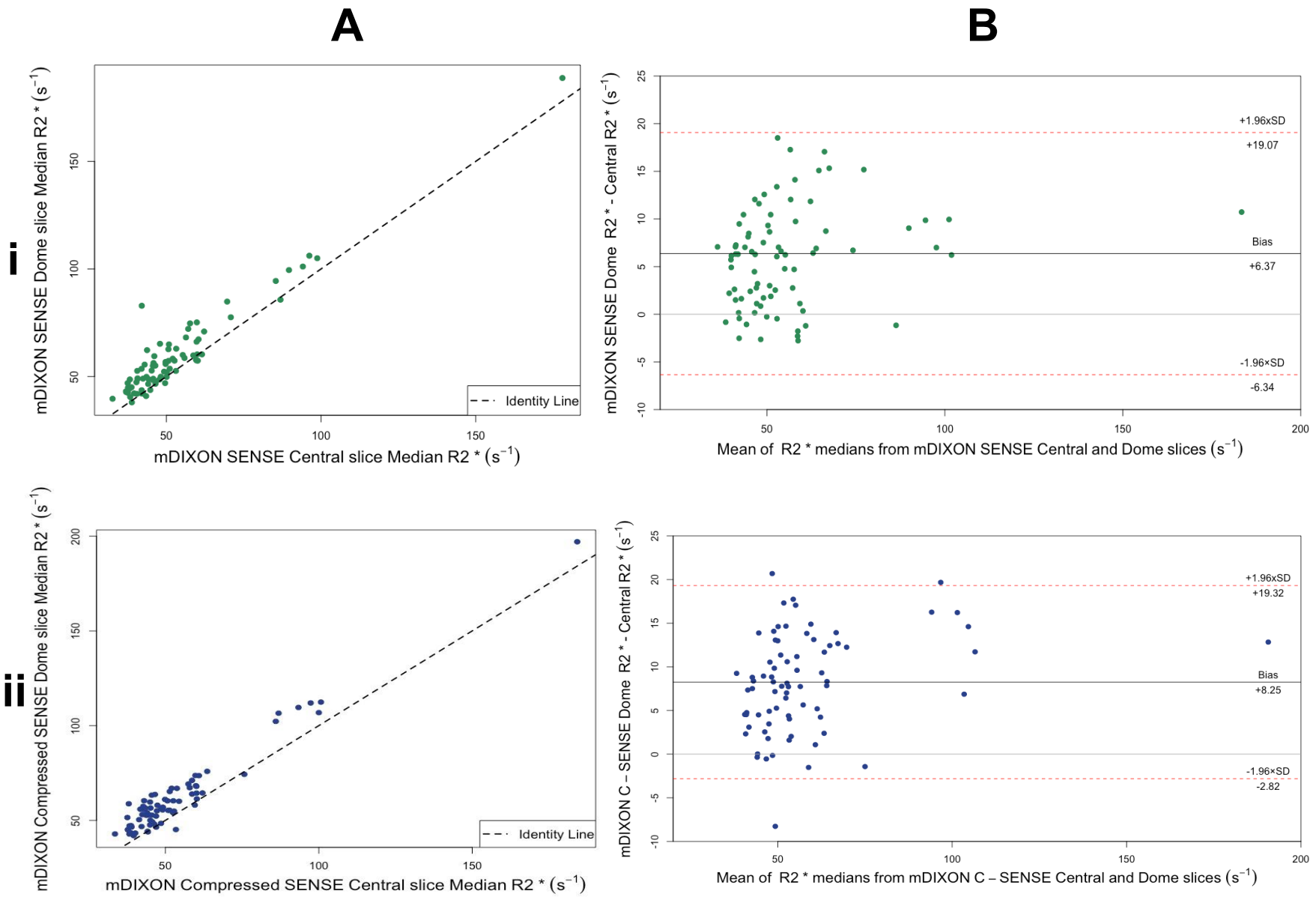


Figure 72 (A) Scatter plot (B) Bland Altman plots, comparing the R_2* median of the segmentations of the central slice and dome slice (i) For mDIXON SENSE and (ii) mDIXON Compressed SENSE

Appendix C Additional Figures from Chapter 8

Comparison of R_2* at different locations in the liver across vendors

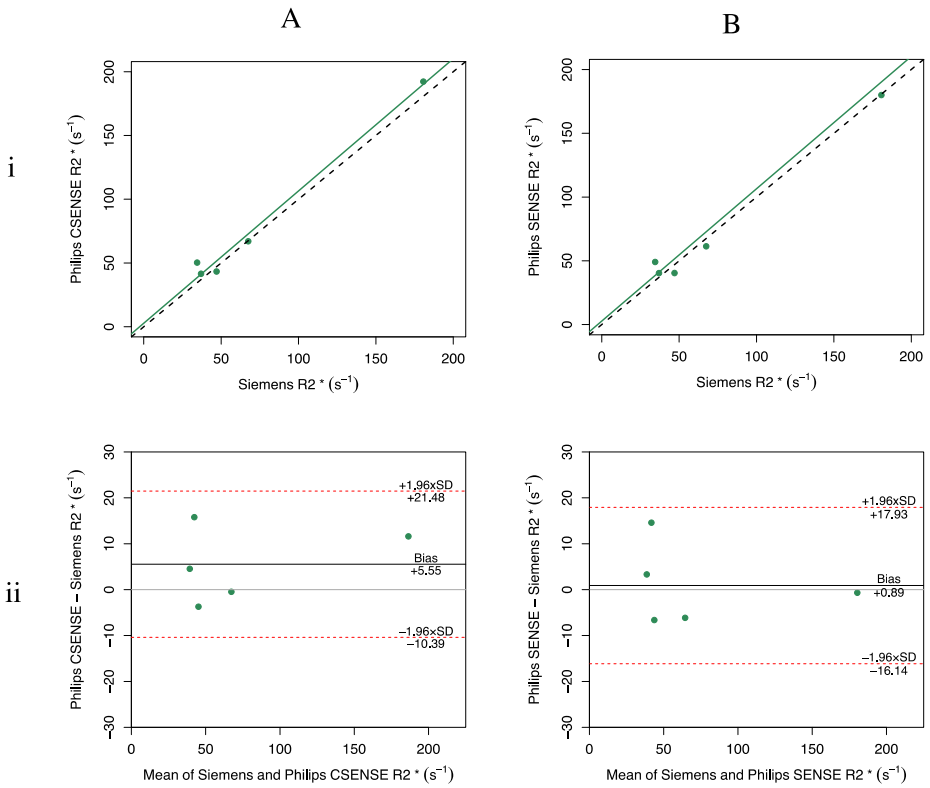


Figure 73 Volunteer upper slice R_2*: (i) Scatter plots and (ii) Bland Altman plots comparing (A) the Philips optimised sequence with Compressed SENSE with the Siemens optimised sequence, (B) the Philips optimised sequence with SENSE with Siemens optimised sequence. On the scatter plot the solid green line is the line of best fit, the black dashed line is the identity line.

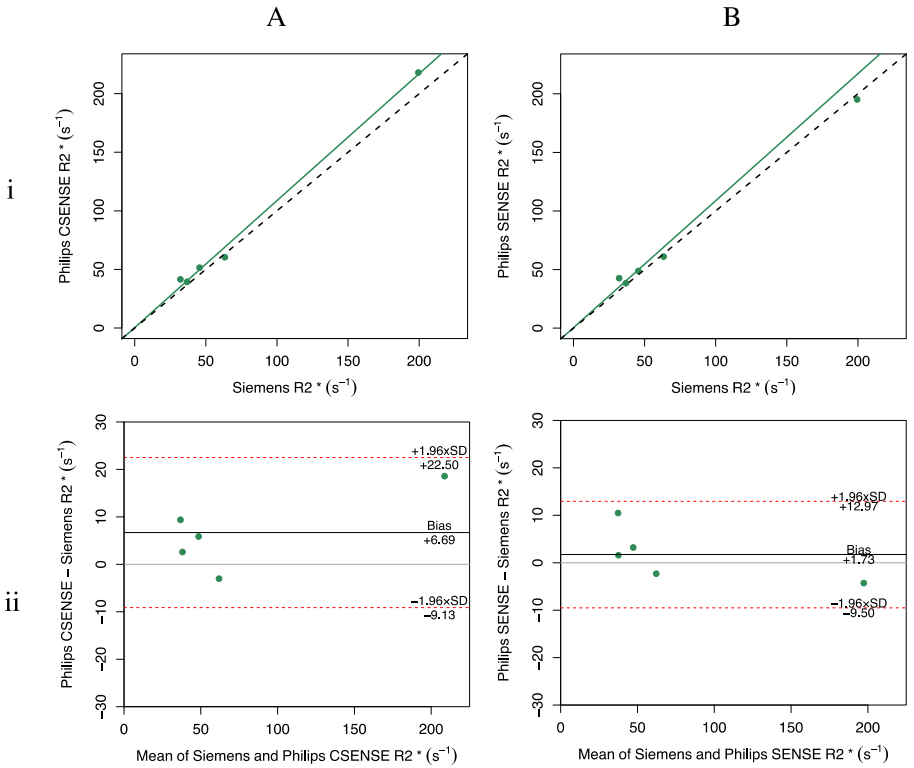


Figure 74 Volunteer lower slice R_2*: (i) Scatter plots and (ii) Bland Altman plots comparing (A) the Philips optimised sequence with Compressed SENSE with the Siemens optimised sequence, (B) the Philips optimised sequence with SENSE with Siemens optimised sequence. On the scatter plot the solid green line is the line of best fit, the black dashed line is the identity line

Appendix D Supplementary Material to Chapter 9

Tube scanning results

The results from the scanning of the tubes are given in Table 47, including the initial scan and the repeated scan.

Concentration	Desired		MOST Initial		MOST repeat	
	Desired R_2* (s^{-1}) (3T)	Equivalent iron using Alam+Hankins mg/g dw	Median R_2* (s^{-1}) (3T)	Equivalent iron using Alam+Hankins mg/g dw	Median R_2* (s^{-1}) (3T)	Equivalent iron using Alam+Hankins mg/g dw
1	37.15	0.4	37.17	0.4	34.43	0.4
2	44.47	0.5	42.86	0.5	42.31	0.5
3	51.79	0.6	49.50	0.6	50.19	0.6
4	59.11	0.7	56.13	0.7	57.84	0.7
5	66.43	0.8	64.13	0.8	61.16	0.7
6	73.75	0.9	70.56	0.9	71.88	0.9
7	81.07	1.0	78.64	1.0	78.98	1.0
8	88.39	1.1	86.46	1.1	94.22	1.2
9	95.71	1.2	93.10	1.2	94.14	1.2
10	103.03	1.3	100.23	1.3	101.45	1.3
11	110.34	1.4	108.91	1.4	110.61	1.4
12	117.66	1.5	113.46	1.4		
13	124.98	1.6	123.61	1.6	125.47	1.6
14	132.30	1.7	131.07	1.7	131.91	1.7
15	139.62	1.8	138.61	1.8	139.67	1.8
16	146.94	1.9	146.72	1.9	144.96	1.9
17	154.26	2.0	154.41	2.0	154.21	2.0

Table 47 Desired R_2* and equivalent iron, Measured R_2* and equivalent iron using MOST, initial, and repeated acquisition results for the tubes

Temperature was not recorded for the initial scan. The repeat scan was 8 weeks later, and the temperature recorded. The placement of the tubes was slightly different, and one of the tubes could no longer be included as it had lost solution and so was mostly air. The temperature was 22.1 °C at the start and 22.0 °C at the end. The scatter and Bland Altman plots comparing the initial MOST results with the results from the repeated acquisition are shown in Figure 75. Spearman's rho was >0.99 ($p < 0.001$) and ICC was >0.99 ($p < 0.001$). The R_2* values and equivalent iron are shown in Table 47. The strong

correlation, excellent agreement and small bias indicate that the results are consistent across both scans, despite the likely temperature difference between the two, and the different positions of the vials. There were some tubes with larger differences, which could have been due to the different positions and changes in concentration if the tube wasn't fully sealed.

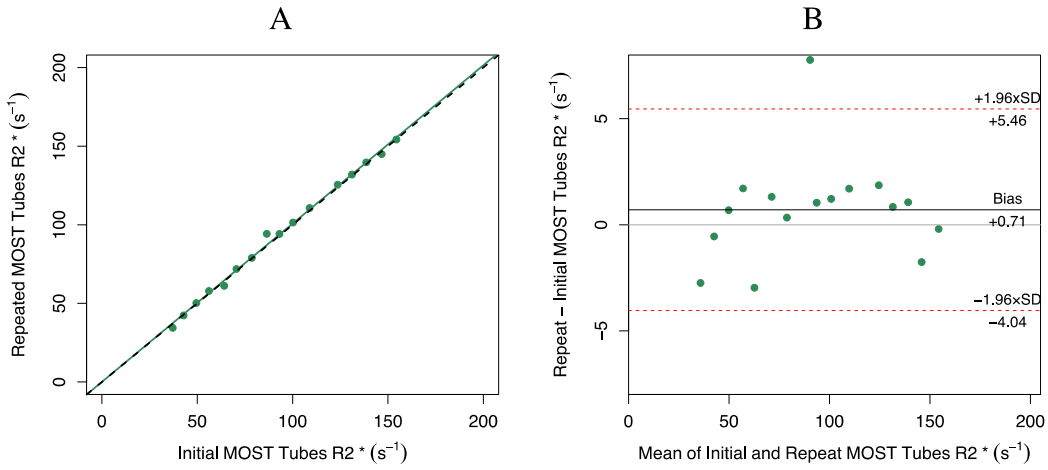


Figure 75 (A) Scatter plot and (B) Bland Altman plot comparing desired R_2* with measured R_2* for the repeated MOST acquisition of the tubes. On the scatter plot the solid green line is the line of best fit, the black dashed line is the identity line.

Comparison of sphere results with reference values

The following figures show the scatter plots and Bland Altman plots comparing the measured ROI values from the spheres with the reference values. Figure 76, Figure 77 and Figure 78 shows the plots for Philips 3 T for large, medium, and small spheres respectively, scanned with SENSE and Compressed SENSE optimised sequences. Figure 79 shows the plots for Siemens for all three sphere sizes.

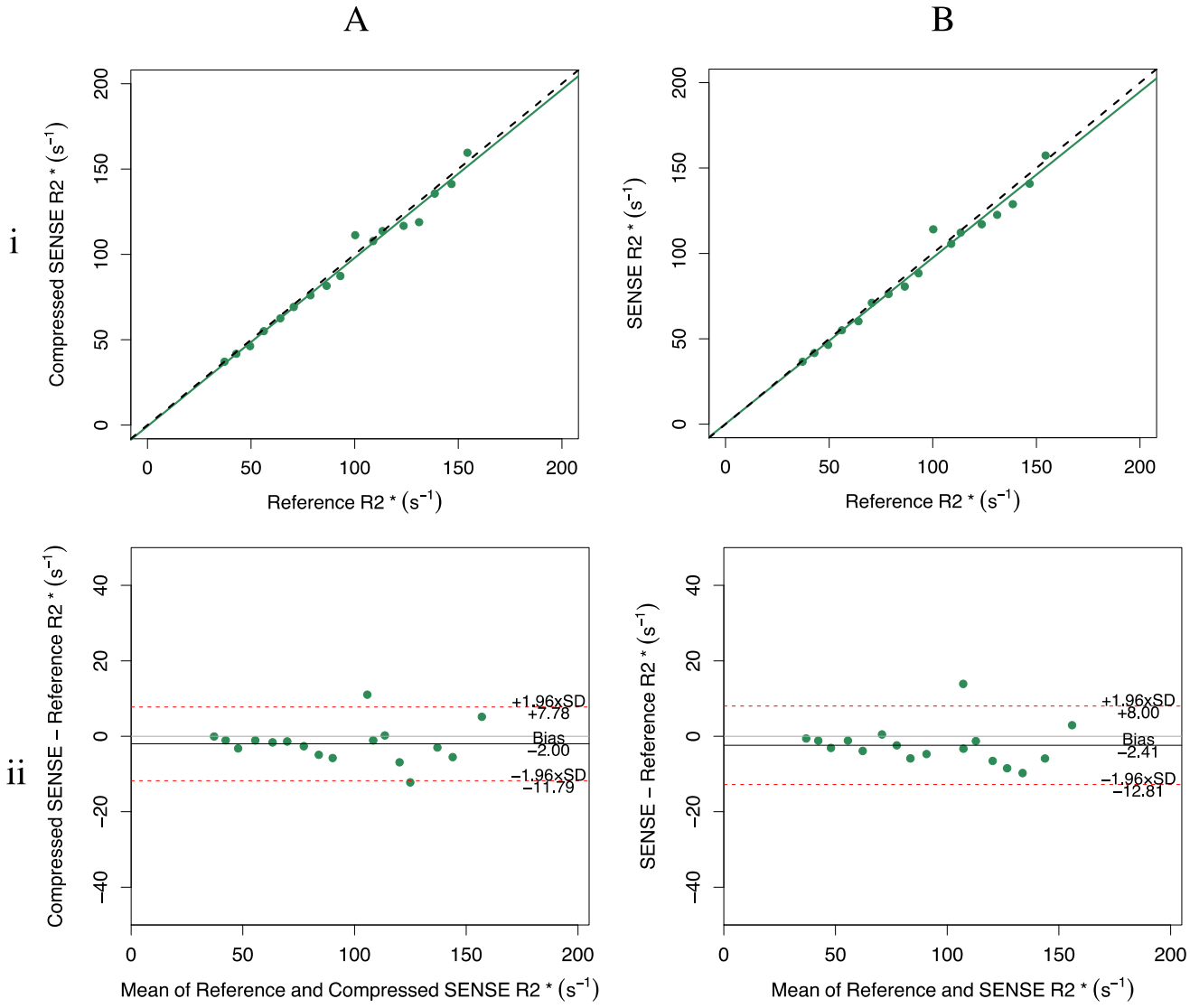


Figure 76 For large spheres only scanned on Philips 3 T: (i) Scatter plots and (ii) Bland Altman plots comparing (A) the optimised sequence with Compressed SENSE with the reference values, (B) the optimised sequence with SENSE with the reference. On the scatter plot the solid green line is the line of best fit, the black dashed line is the identity line.

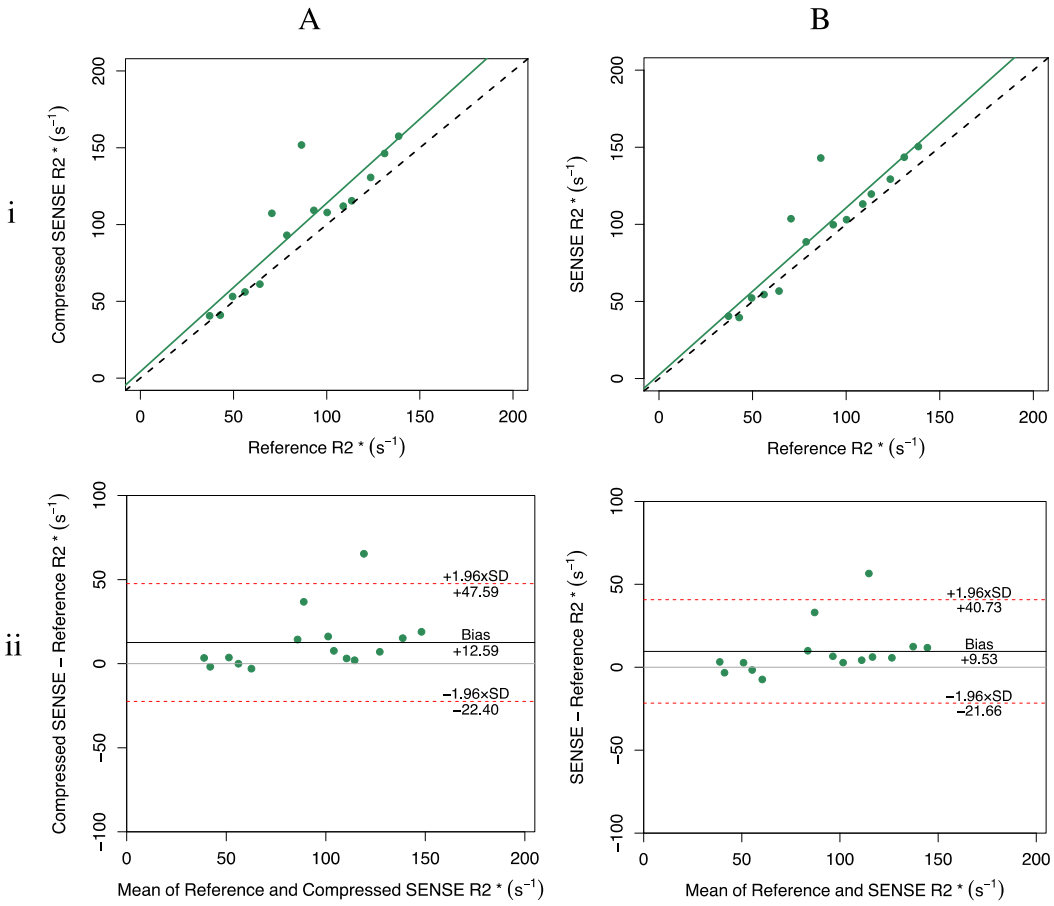


Figure 77 For medium spheres only scanned on Philips 3 T: (i) Scatter plots and (ii) Bland Altman plots comparing (A) the optimised sequence with Compressed SENSE with the reference values, (B) the optimised sequence with SENSE with the reference. On the scatter plot the solid green line is the line of best fit, the black dashed line is the identity line.

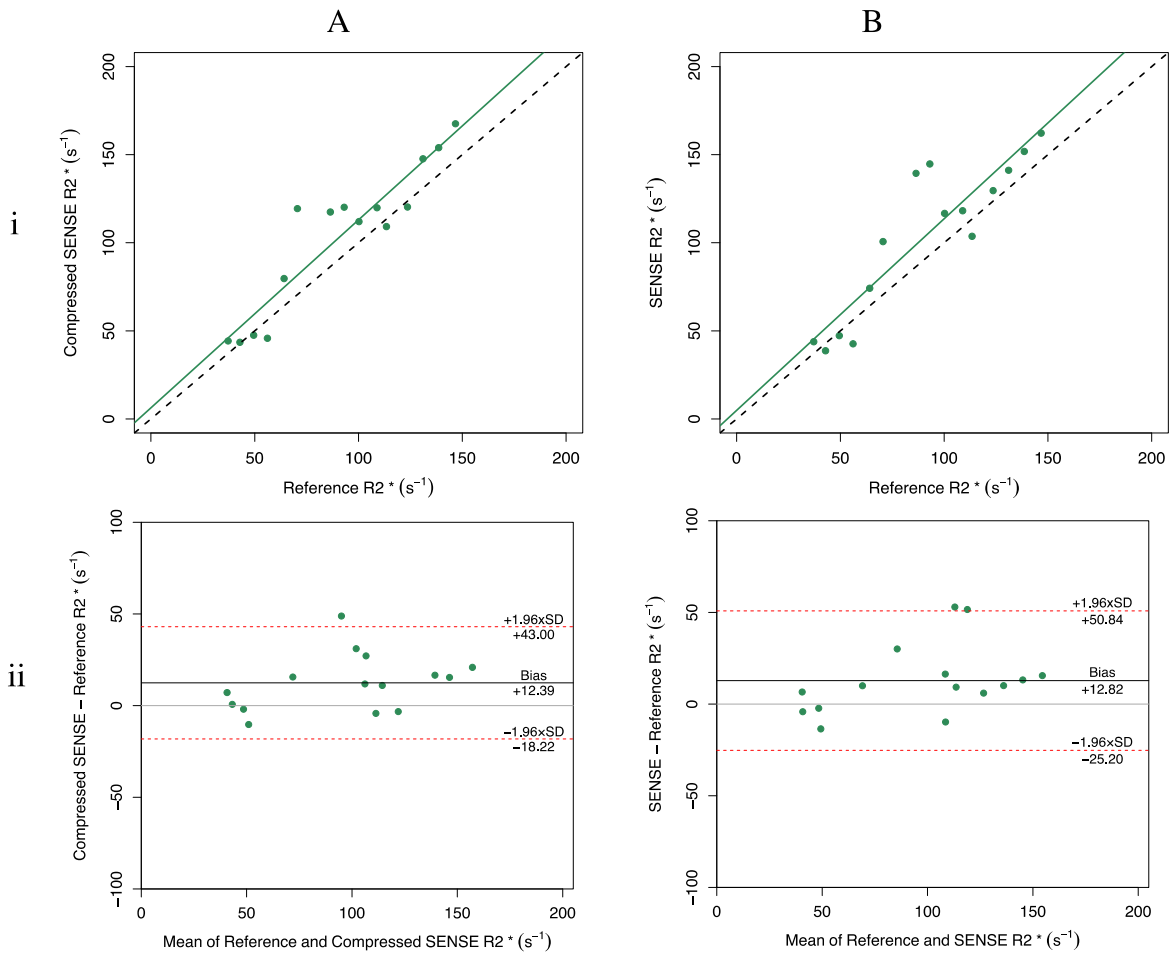


Figure 78 For small spheres only scanned on Philips 3 T: (i) Scatter plots and (ii) Bland Altman plots comparing (A) the optimised sequence with Compressed SENSE with the reference values, (B) the optimised sequence with SENSE with the reference. On the scatter plot the solid green line is the line of best fit, the black dashed line is the identity line.

i

ii

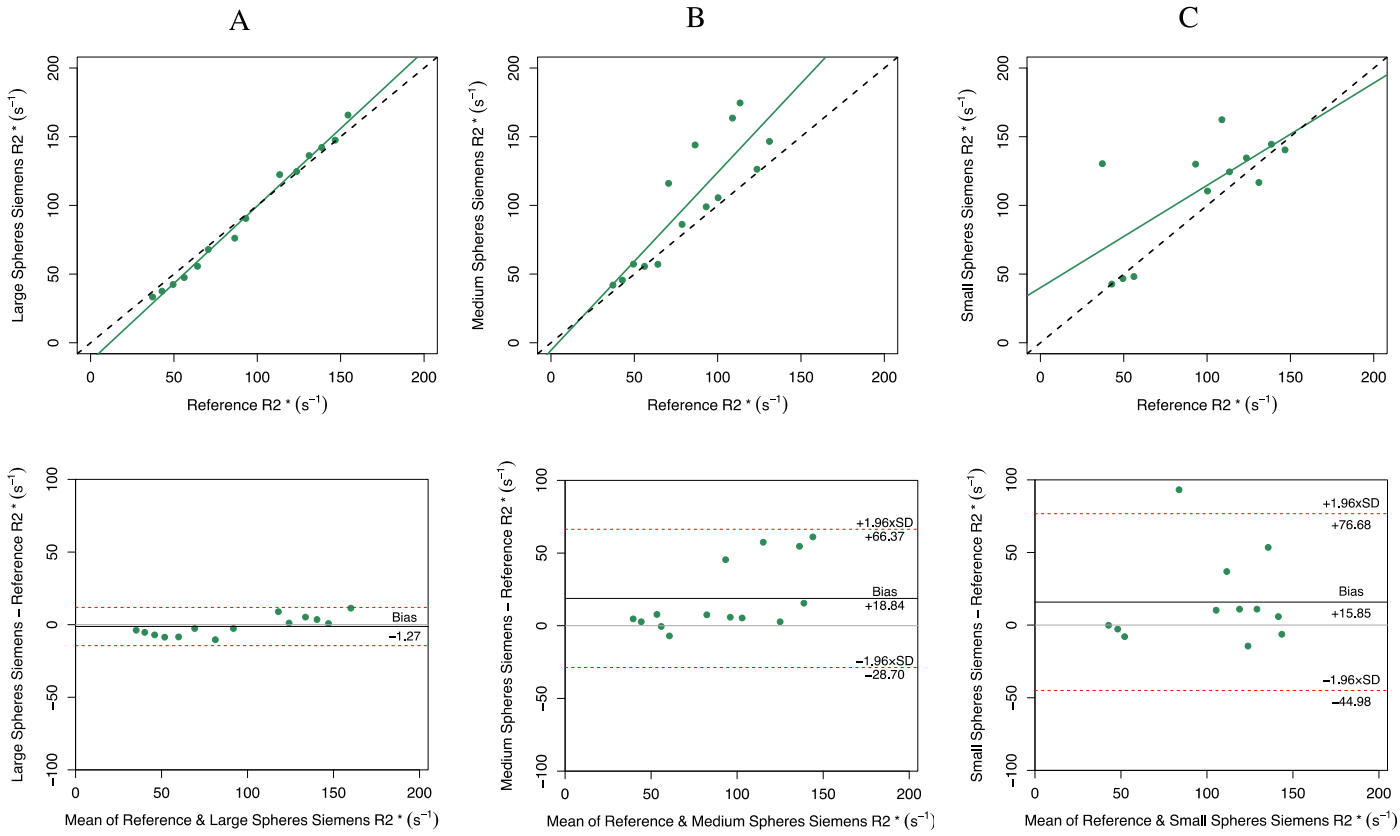


Figure 79 (i) Scatter plots and (ii) Bland Altman plots comparing the results for (A) the large spheres, (B) medium spheres and (C) small spheres scanned with Siemens optimised sequence, with the reference values. On the scatter plot the solid green line is the line of best fit, the black dashed line is the identity line. For (A) (ii) the limits of agreement were +11.88 s^{-1} and -14.42 s^{-1}.

Comparison between Philips and Siemens

To compare R_2* results from the Philips and Siemens acquisitions, Spearman's rho and ICC were calculated (Table 48) and scatter plots and Bland Altman plots made (Figure 80 and Figure 81). Only the large and medium sphere results were compared due to the limited number of small spheres in the Siemens analysis.

The results show that there was strong correlation and agreement between the Philips and Siemens results, although they were slightly weaker for medium compared to large spheres. The bias was also small (less than 7.33 s^{-1}), for both large and medium spheres, indicating that the sequences on both vendors do give similar measurements. However, the LoA were >7.33 s^{-1} and were increasingly negative for higher R_2* i.e., both Philips sequences give increasingly lower values compared to Siemens for higher R_2*. This was more evident for the larger spheres, as the medium spheres with higher R_2* were excluded.

The small negative bias may be due to the lower temperature of Philips towards the end of the scan (at the end of the scan it was 20.6 °C compared to 23.6 °C at the end of the Siemens scan). A previous study found 1.6 %/°C increases in T_2 for $MnCl_2$[108]. Philips R_2* was increasing lower than Siemens for higher R_2*, i.e., would give a lower iron reading compared to Siemens. This is unlikely to be due to any temperature differences as the plate with the highest R_2* was scanned first for Philips, when the temperature was highest (23.5 °C) and thus was similar to the Siemens temperatures (which were 22.9 °C at the start and 23.6 °C at the end). A reason could be that the initial TE is earlier in Philips (1.00 ms compared to 1.45 ms), making higher R_2* easier to estimate (as having an earlier first TE means more signal is measured before it decays away).

When comparing with the *in vivo* results from Chapter 8 is interesting to note that for the phantom there was a negative bias i.e., Philips gave lower results, whereas for the healthy volunteers there was a positive bias, i.e., Philips gave higher results. It may be that differences in temperature between the two vendors influenced the phantom results. The LoA were generally slightly wider for the *in vivo* comparison compared to the results from the large spheres, but smaller than for the medium spheres. Unlike in the phantom, no increasing negative difference between Philips and Siemens was seen in the case with higher R_2*.

Sphere size	Comparison	Spearman's rho (p value)	ICC (p value)
Large spheres	Siemens 3 T Optimised sequence and Philips 3 T Optimised Sequence with Compressed SENSE	>0.99 (p <0.001)	0.99 (p <0.001)
	Siemens 3 T Optimised sequence and Philips 3 T Optimised Sequence with SENSE	1.0 (p <0.001)	0.99 (p <0.001)
Medium spheres	Siemens 3 T Optimised sequence and Philips 3 T Optimised Sequence with Compressed SENSE	0.92 (p <0.001)	0.87 (p <0.001)
	Siemens 3 T Optimised sequence and Philips 3 T Optimised Sequence with SENSE	0.93 (p <0.001)	0.89 (p <0.001)

Table 48 Spearman's rho and ICC for R_2* of Siemens compared to Philips sequences

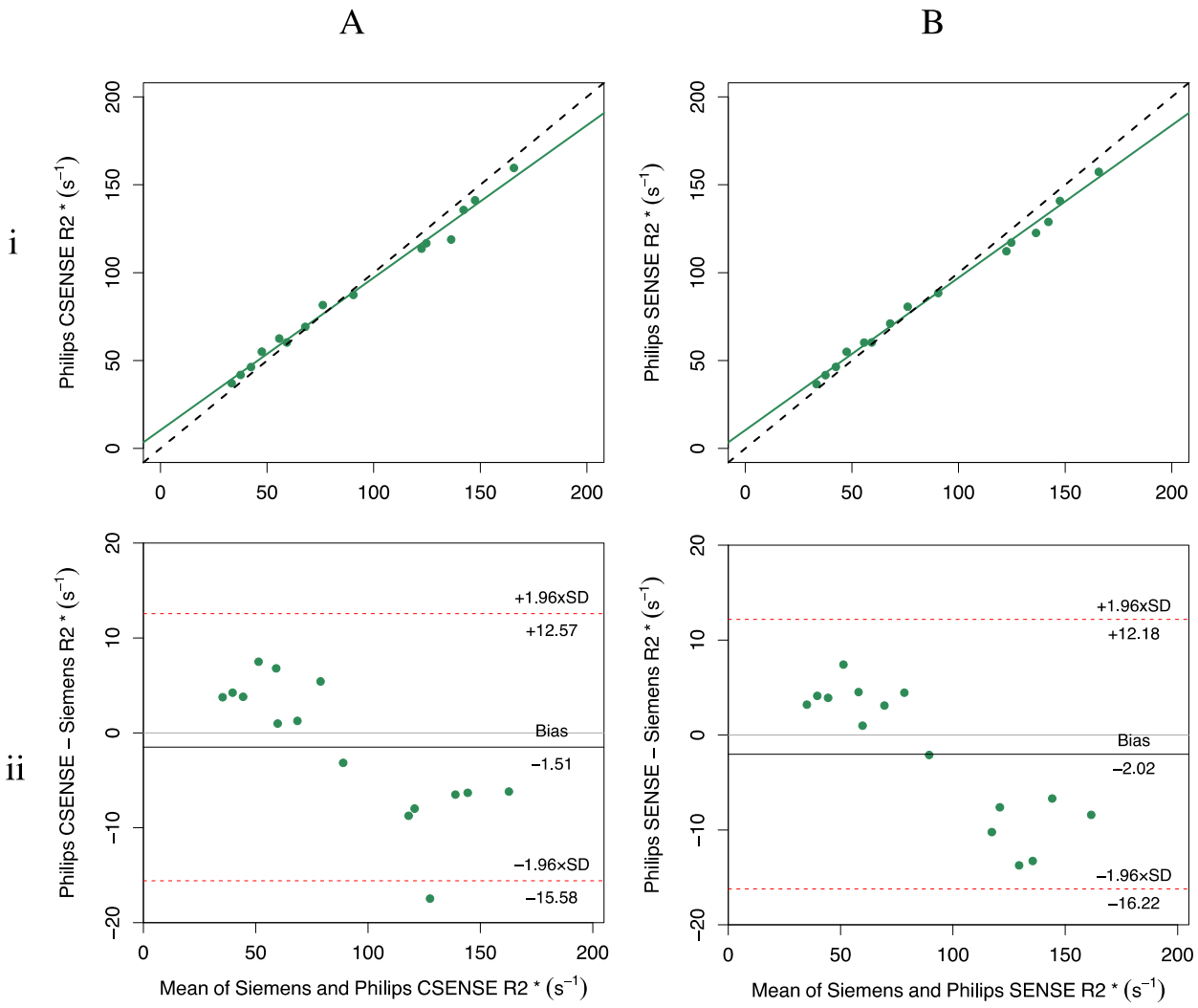


Figure 80 Using measurements from the large spheres only, (i) Scatter plots and (ii) Bland Altman plots comparing (A) the Philips optimised sequence with Compressed SENSE with the Siemens optimised sequence, (B) the Philips optimised sequence with SENSE with Siemens optimised sequence. On the scatter plot the solid green line is the line of best fit, the black dashed line is the identity line.

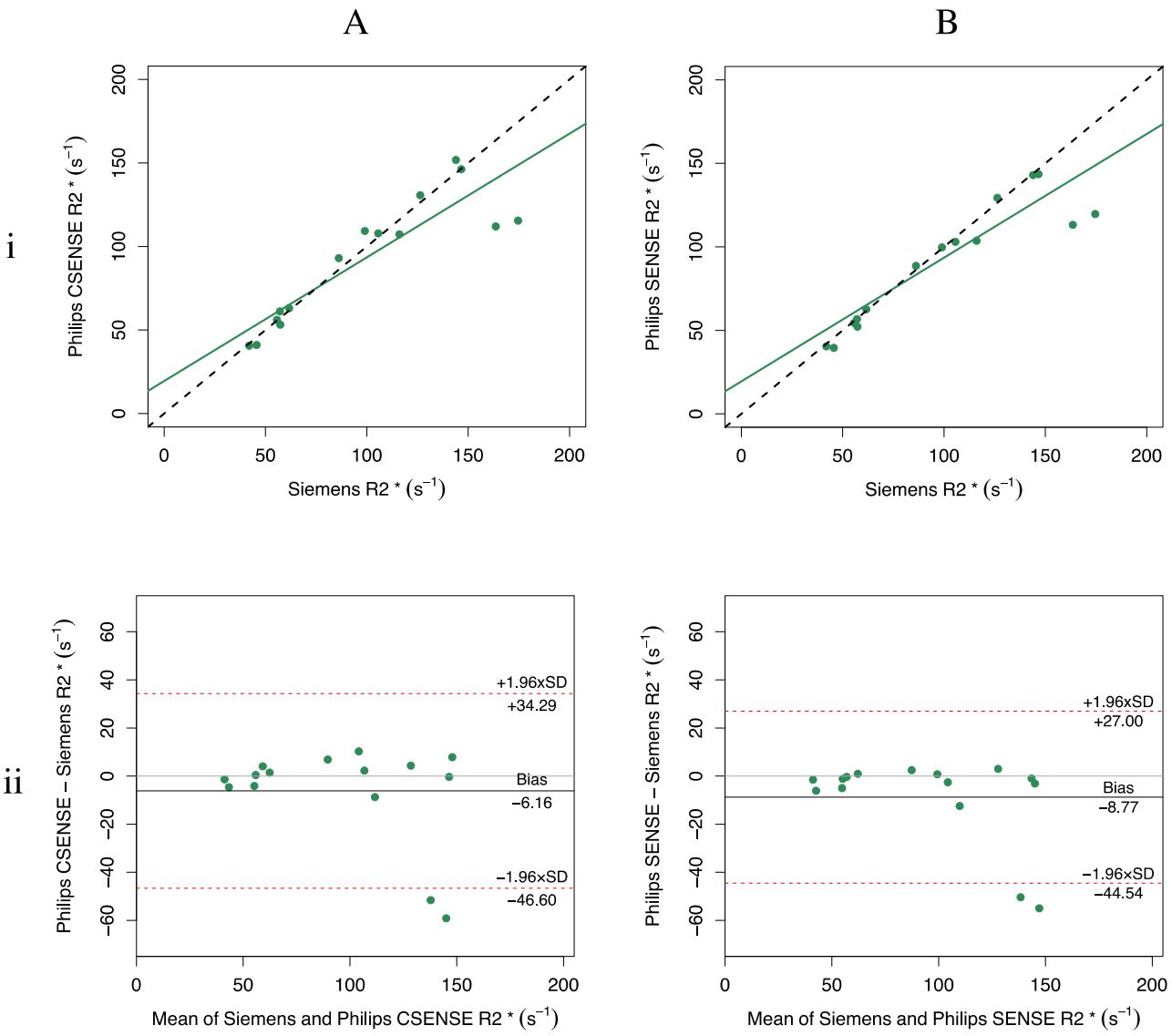


Figure 81 Using measurements from the medium spheres only, (i) Scatter plots and (ii) Bland Altman plots comparing (A) the Philips optimised sequence with Compressed SENSE with the Siemens optimised sequence, (B) the Philips optimised sequence with SENSE with Siemens optimised sequence. On the scatter plot the solid green line is the line of best fit, the black dashed line is the identity line.