5-2-2023

Simulation of a Virtual Iron-Overload Model and R2* estimation using Multispectral Fat-Water Models for GRE and UTE Acquisitions using MRI

Prasiddhi Neupane

Simulation of a Virtual Iron-Overload Model and R2* estimation using Multispectral Fat-Water Models for GRE and UTE Acquisitions using MRI

by

Prasiddhi Neupane

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Major: Biomedical Engineering

The University of Memphis

May 2023

Acknowledgements

I would not be where I am today without the guidance and friendship of so many people who have helped me grow as a student, researcher, and engineer. First and foremost, I would like to thank my advisor Dr. Aaryani Tipirneni-Sajja, for consistently guiding and supporting me throughout the past two years and for instilling in me the yearning for research. Her work ethic is something I have always admired, and I hope to inherit some of it. I am also grateful to Dr. Amy Curry, Dr. Herickhoff, and Dr. Eddie Jacobs for their insightful recommendations throughout my thesis preparation.

I owe my deepest gratitude to my lab members, all of whom have played important roles in my graduate school life. First of all, I would like to thank Utsav Shrestha for helping me get started with my thesis project and for always lending a helping hand. His knowledge of computational modeling and MRI was instrumental in kindling my interest in MRI. I would also like to thank Sarah Brasher, without whom my graduate school experience would not have been as enjoyable. Her constant companionship throughout all of the classes we took and our quick lunch and Starbucks breaks will be a treasured memory for me. I would also like to thank Juan Esparza and Hayden Johnson for always being warm and welcoming when I needed it. Additionally, I would like to thank Dr. Erno Lindner, whose coffee I believe has fueled this entire journey. I will miss making coffee for him every day, and wherever I see Folgers I will think of him.

I would like to thank my father, Prakash Neupane, for always encouraging me to strive forward, and my mother, Kanchan Neupane, for her unending love and sacrifices. I would like to thank my dear grandmother, Rama Devi Neupane, for being the warmest and most loving person

in my life. I would also like to thank everyone else in my large and loving family, especially my siblings and cousins- Dipesh, Dikshya, Yogyata, Pragyee, Riddhi, Pranav, and Shuvanga. Everything I do is for you, and I can only hope that all of this will one day allow us to live near each other happily ever after. Thank you.

Preface

Chapter 4, the main body of this thesis, is composed of the following manuscript:

"Simulation of a Virtual Liver Iron-Overload Model and R2* Estimation Using Multispectral Fat-Water Models for GRE and UTE Acquisitions", submitted to *NMR in Biomedicine* and under review.

Abstract

Iron overload is an excessive accumulation of iron in the body and can be either inherited or acquired through chronic blood transfusions. Assessment of hepatic iron concentration (HIC) is important in the management and monitoring of iron overload. Despite liver biopsy being the gold standard method for assessing HIC, it is invasive, painful, unsuitable for repeated measurements, and carries the risk of bleeding and infection. Magnetic Resonance Imaging (MRI) methods based on transverse relaxation rate (R2*) have emerged as a non-invasive alternative to liver biopsy for assessing HIC. Multispectral fat-water-R2* modeling techniques, such as the non-linear square (NLSQ) fitting and autoregressive moving average (ARMA) models, have been proposed to provide more accurate assessments of iron overload by accounting for the presence of fat, which can otherwise confound R2*-based HIC measurements in conditions of co-existing iron overload and steatosis. However, the R2* estimation by these multispectral models has not been systematically investigated for various acquisition methods like the multiecho gradient echo (GRE) and ultrashort echo time (UTE) across the full clinically relevant range of HICs. To address this challenge, a Monte Carlo-based iron overload model based on true iron morphometry and histological data was constructed, and MRI signals were synthesized at 1.5 T and 3 T field strengths. This study compared the accuracy and precision of multispectral NLSQ and ARMA models against the monoexponential model and published in vivo R2*-HIC calibrations in estimating R2*. The results showed that, for GRE acquisitions, ARMA and NLSQ models produced higher slopes compared to the monoexponential model and published in vivo R2*-HIC calibrations. However, for UTE acquisitions for shorter echo spacing ($\leq$ 0.5 ms) and longer maximum echo time, TEmax ($\geq$ 6 ms), both multispectral and monoexponential signal models produced similar R2*-HIC slopes and precision values across

the full clinical spectrum of HICs at both 1.5 T and 3 T. The results from the simulation studies were validated using phantoms and patient data. Future work should investigate the performance of multispectral models by simulating liver models in coexisting conditions of iron overload and steatosis to investigate simultaneous and accurate quantification of both R2* and fat.

Table of Contents

Section **Page #**

List of Tables

List of Figures

Abbreviations

HIC	Hepatic Iron Concentration
R2*	Effective transverse relaxation rate
GRE	Gradient echo sequence
NLSQ	Nonlinear square fitting
ARMA	Autoregressive moving average
FF	Fat fraction
UTE	Ultrashort Echo Time
TE_1	First echo time
ΔTE	Echo spacing
TEmax	Maximum echo time
3D	Three dimensional
CoV	Coefficient of variation
SNR	Signal to noise ratio
Mono-exp	Monoexponential
SD	Standard Deviation

Chapter 1 – Introduction

Significance

Iron overload is a medical condition characterized by excessive accumulation of iron in the body, which can result from inherited genetic mutations or from treatments for certain diseases such as sickle cell anemia, beta-thalassemia, or cancer.[1,2] Iron overload can cause various health problems, including liver disease, cardiomyopathy, joint pain, diabetes, and skin hyperpigmentation. Since the liver plays a key role in regulating iron homeostasis, it is often the organ most affected by iron overload, and excess iron in the liver can lead to chronic liver disease, cirrhosis, fibrosis, and hepatocellular carcinoma.[3]

Liver iron stores increase proportionally with total body iron.[4] Therefore, assessing the amount of liver iron is essential for effective treatment and management of iron overload. Traditionally, liver biopsy has been used to measure the amount of iron in the liver, but this procedure is invasive and carries risks of complications such as infection and bleeding. In recent years, magnetic resonance imaging (MRI) techniques based on effective transverse relaxation rate (R2*) have emerged as a non-invasive alternative to liver biopsy for estimating hepatic iron concentration (HIC) using clinically accepted R2* vs biopsy HIC calibration curves.[5-7] Previous in vivo calibration studies have utilized multiecho gradient echo (GRE) imaging and monoexponential fitting model to estimate R2*.[5,6] However, in the presence of concomitant iron and fat in the liver, the MRI signals do not follow a pure monoexponential decay due to oscillations induced by the fat signal.[8] This limits the accuracy of existing MRI techniques that use monoexponential fitting in estimating HIC accurately since fat accumulation is a common condition, and patients suspected of iron overload may have presence of hepatic fat as well. To address this challenge, researchers have developed multispectral fat-water-R2* modeling

techniques such as non-linear square (NLSQ) fitting and autoregressive moving average (ARMA) that can account for the confounding effects of both iron and fat on MRI signals. Both NLSQ and ARMA signal models have been validated to perform well in mild and moderate iron overload, however they have been shown to fail in high iron overload conditions due to rapid signal decay that cannot be adequately captured using the current GRE acquisitions with TE_1 ~1 ms.[8,9] Alternatively, multiecho ultrashort echo time (UTE) imaging, with TE_1 ~0.1 ms, has been recently shown to improve the accuracy and precision of R2* estimation over a wider clinical range of HICs than conventional GRE acquisitions.[10,11] Multispectral models have not yet been systematically investigated for UTE acquisitions and evaluated for accuracy and precision across the entire range of HICs. Kee et al. used NLSQ model for R2* quantification using a 3D UTE cones acquisition at 3T in one study, but this study only used mild to moderate iron overload patients.[12] In this project, we aim to perform a thorough and systematic investigation of multispectral models in estimating R2* as well as FF under multiple GRE and UTE acquisitions throughout the full clinically relevant range of HIC. However, evaluating these models using multiple MRI acquisitions in phantoms or patients can be costly and challenging. Monte Carlo simulations, on the other hand, have been demonstrated in previous studies to produce realistic virtual liver iron overload models and synthesize MRI signals resembling in vivo behavior.[13-15]

Therefore, for this study, we propose to construct Monte Carlo-based virtual liver models with varying iron overload conditions, synthesize MRI signals for different GRE and UTE acquisitions for varying sets of echo spacings (ΔTE) and maximum echo time (TEmax) at magnetic field strengths of 1.5 T and 3 T, and estimate R2* using multispectral fat-water-R2* models. The goal of this study is to assess the accuracy and precision of multispectral fat-water-R2* models in estimating R2* and FF under different GRE and UTE acquisitions and to validate

their results against reference monoexponential and published R2*-HIC calibrations. We also aim to validate the results from our simulation study using MRI images of phantoms and patients under multiple iron overload conditions. We believe that our study will be beneficial in the investigation and validation of quantitative MRI approaches for evaluation of iron overload and will contribute to the development of MRI as a non-invasive diagnostic tool for iron overload assessments.

Objectives

The main objectives of our study are to construct virtual liver models with varying HICs using Monte Carlo simulations for evaluating the performances of multispectral models for various GRE and UTE acquisitions and to validate the results using phantom and in vivo MRI data. The following are the specific aims of our research:

Aim 1: Develop Monte Carlo-based iron overload model and simulate MRI signals to evaluate performance of fat-water-R2* signal models.

Aim 2: Investigate and validate the results derived from simulation study, by utilizing MRI images of both phantoms and patients.

Outline of the Thesis

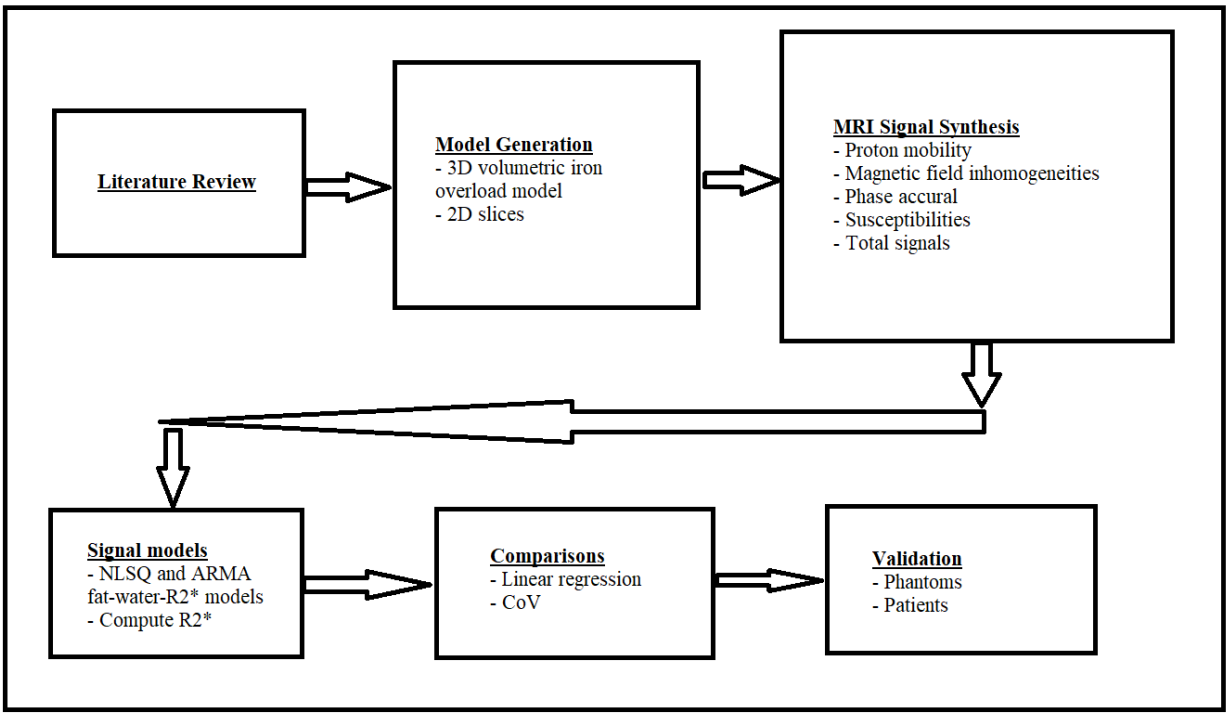

Figure 1. An outline of the thesis.

The remainder of thesis is organized as follows:

Chapter 2 discusses about iron homeostasis in the body, pathophysiology of iron overload, treatment and management, and current available methods for assessment. Additionally, MRI principles are introduced, and the chapter describes how MRI techniques are utilized to assess iron overload, including the limitations of current MRI acquisitions and signal models.

Chapter 3 outlines the development and implementation of a Monte-Carlo based iron overload model, which generates MRI signals. The chapter also describes the implementation of different signal models for predicting R2*.

Chapter 4 describes the methodology and results of a systematic comparison between the accuracy and precison of R2* and FF estimated by the different signal models using the Monte Carlo based iron overload simulations.

Chapter 5 describes the methodology and results of a validation approach of the simulation results, using phantoms and patients.

Chapter 6 concludes the thesis by discussing the main findings of the simulations, phantom, and patient studies, and presents some prospects for future research.

Chapter 2 - Literature Review

Iron Homeostasis and Overload

Iron overload is the excessive accumulation of iron in the body. Because the liver plays a key role in maintaining iron homeostasis, it is the organ most commonly impacted by iron overload. The heart and endocrine organs like pituitary gland, pancreas, gonads, etc. are among the other organs that can be affected by iron overload.[1,4] Excess iron in the liver can progress to chronic liver disease, cirrhosis, fibrosis, and, eventually, liver failure.[3] Other possible manifestations of iron overload are cardiomyopathy, arthropathy, cancers, diabetes, skin hyperpigmentation, etc.[1,16] It is estimated that 16 million Americans suffer from some form of iron overload.[1] Iron overload is more common in Europe, Australia, and other western countries where there are larger populations of people of Celtic origin.[1] Further, white Americans are more likely to develop iron overload than black Americans.[1] Men have been discovered to have a higher risk of developing iron overload than women due to the fact that women lose iron from blood during their monthly menstrual cycle.[1] Patients of iron overload are often asymptomatic until they reach adulthood, but extreme fatigue, lethargy, and arthralgias are the most prevalent symptoms.

The human body contains approximately 50 mg of iron per kg for men and 40 mg of iron per kg for women.[17] Most of the iron in the body is bound to hemoglobin in red blood cells and the remainder is stored in different forms such as ferritin, hemosiderin, and transferrin.[16-18] Dietary iron is absorbed in the duodenum and transported by ferroportin to the bloodstream, where it binds to transferrin.[19] Transferrin then transports the iron to the bone marrow for erythropoiesis and recycling, as well as to the liver and other organs for storage. Ferritin is a protein that stores iron in a stable form inside a hollow cage-like structure to prevent the oxidation of free iron and the production of harmful reactive oxygen species. When ferritin is

saturated with iron, excess iron can be stored in hemosiderin. Furthermore, iron can be excreted out by the liver via bile and be transported into the small intestine, where it can undergo enterohepatic circulation, or can be eliminated from the body via feces.[1,3] The regulation of iron homeostasis in the body is controlled by hepcidin, a protein produced by the liver. Hepcidin reduces the number of ferroportin transporters in enterocyte cell membranes, lowering the amount of iron absorbed from food.[18] Inflammatory cytokines, lipopolysaccharides, and the HFe protein stimulate hepcidin production, which also reduces the number of ferroportins in liver and spleen cell membranes to inhibit the release of iron into the bloodstream.[20]

Iron overload is of two types: primary and secondary. Primary iron overload, also known as hereditary hemochromatosis, is an inherited genetic condition of a defective HFe gene.[16] HFe gene, located on the short arm of chromosome 6, is responsible for the production of HFe proteins involved in sensing and regulating iron levels in the body and determining iron storage.[1] Secondary hemochromatosis develops as a result of chronic blood transfusions used to treat hematologic diseases such as sickle cell disease, beta-thalassemia or as part of cancer treatment.[8,21] When there is excessive amount of red blood cells (RBCs) in the blood, the body absorbs excess iron by all the hemoglobin in the RBCs.[1] Or when RBCs are fragile, like in the case of anemia, they die more easily, and when iron is recycled, it results in iron deposition in the body.[1] Secondary hemochromatosis is also associated with the presence of underlying diseases such as thalassemia and spherocytosis.[21]

A normal hepatic iron concentration (HIC) is considered approximately 0.2-2 milligrams iron per gram of dry weight of liver.[17,22] Because our body lacks an active control mechanism to regulate iron levels, except for the passive elimination via urine and other waste products and menstrual cycles in women, excess iron caused by all the different types of hemochromatosis is

initially deposited in ferritin and hemosiderin.[21] When ferritin and hemosiderin get saturated, the free iron begins to deposit in various organs, especially liver- the main organ for iron storage. Excess iron can deposit within hepatocytes, sinusoidal and portal macrophages, sinusoidal and portal endothelial cells, and biliary (ductal) cells, depending on the type of iron overload; the excess iron from intestinal absorption comes into the liver lobules through the portal vein and deposits within hepatocytes as fine granules.[23] Excess iron can also be deposited in the Kupffer cells and portal macrophages, or in all the above cells causing massive iron overloading. Functional iron of the body, located in hemoglobin, myoglobin, iron-containing enzymes, and storage proteins ferritin and hemosiderin, remains unaffected and efficient, but the excess iron deposition is toxic and causes a variety of problems.[21]

Treatment and Management

Treatment and management of iron overload depends on the type and severity of the disease. For primary hemochromatosis, the go-to treatment method is phlebotomy.[1,17] Patients must have 500 mL blood drawn out periodically for the rest of their lives, progressively less frequently as iron concentrations return to normal. The aim is to reduce red blood cells, or hemoglobin, from the blood and lower iron level.[17] Secondary hemochromatosis patients, on the other hand, cannot be treated with phlebotomy and must undergo chelation therapy, which is the administration of chelation agents to remove metals (such as iron) from the body.[24] Deferasirox, deferiprone, and deferoxamine are the most commonly used chelation agents for iron overload.[17] If the organs have already been severely affected, like in fibrosis and cirrhosis, the only treatment method is organ transplantation, which comes with its own set of risks.[1] However, if iron overload is detected early enough, it can be reversed, and iron levels can be returned to normal.

Assessment of Hepatic Iron Overload

Assessment of iron overload is an important part of treatment and management. Further, assessment helps in determining the type of iron overload and hence direct the form of therapy (phlebotomy and chelation).[21] Liver iron stores increase proportionally with total body iron, so assessing hepatic iron concentration, abbreviated as HIC, is an effective way of monitoring iron overload for treatment management.[4] Traditionally, HIC is assessed using liver biopsy, which is considered the gold standard for diagnosing liver disease. Liver biopsies can help quantify HIC and assess liver damage; however, they are invasive, painful, and have risks of infection and bleeding.[5,6] As a result, noninvasive techniques for assessing iron overload are currently of great interest and are being extensively researched to replace biopsy.

A noninvasive technique for measuring serum iron, ferritin, or transferrin is an easy and fast way to determine HIC. However, serum iron can be influenced by a variety of factors such as transfusions, inflammation, alcohol consumption and so on.[16] So, despite its simplicity, this method is not as reliable as biopsies.[16,17] Ultrasound is not able to quantify HIC; however, it can assess the extent of organ damage due to iron deposition. Similarly, Computed Tomography (CT) can also detect liver iron qualitatively; attenuation, or x-ray absorption, in liver is higher in iron overload. However, attenuation in liver is not specific to increase with iron content and can further be confounded by the presence of fat and other factors.[21] In addition, CT is not useful for iron overload assessment because repeated CTs are harmful to our bodies and iron overload requires multiple assessments on a regular basis. The most reliable noninvasive technique that has gradually started to be accepted for assessment of iron overload is based on magnetic susceptibility of tissues. Magnetic susceptibility is a dimensionless quantity that indicates the degree of magnetization of a material in response to an applied magnetic field. Liver iron

susceptometry utilizes the difference in susceptibility between normal tissue which is diamagnetic in comparison to iron overload tissue that is paramagnetic.[25] The device that can measure iron concentration, based on the difference in magnetic fields generated in response to an external magnetic field, is called a superconducting quantum interference device (SQUID) and its HIC quantification has been shown to exhibit excellent correlation with HIC determined by biopsy. However, because of its high cost and complexity, there are only four of these devices in the world and as a result, SQUID has not been used commonly in clinical field.[2] Magnetic Resonance Imaging (MRI) also utilizes this concept of susceptibility for the assessment of iron overload, and over the last decade MRI has emerged as a clinically accepted method for quantifying HIC.[6,8]

MRI based Iron Quantification

A primary magnet, gradient magnets, radiofrequency (RF) coil, and the computer system make up an MRI machine. The magnetic strength of commonly used magnets in clinical MRI machines is 1.5 or 3 Tesla (T). Our body is composed of 70% water; in this way hydrogen atoms are present in all tissues and organs of the body. These H atoms each consist of one rapidly spinning proton in their nucleus, and one electron in orbit. The spin of the proton generates a magnetic field, called the magnetic moment. In normal conditions, all the H protons in our body are randomly oriented and hence there is no overall magnetic field due to the protons. On the other hand, protons orient and precess in the same direction as the magnetic field of the primary magnet (longitudinal z-direction) when the body is inside an MRI machine.[16] At this point, if the RF coil in the MRI sends a wave of RF signals, then the protons get tipped off and start precessing in phase in the transverse direction (xy-direction). According to the famous Lenz law,

when a wire is placed inside a rapidly changing magnetic flux, it induces an electrical signal in the wire. Similarly, in MRI, due to rapidly changing magnetic flux from the tipping of the protons, a signal is generated in the RF coil, which is detected by RF coils in the MRI machine. The signal generated does not last forever; the protons in transverse plane go from in-phase to out of phase and go back to realigning with the magnetic field in the z direction. This process of knocking spins repeatedly until enough signals are detected by the scanners is what produces the MRI image of different parts of the body.

The time taken by the tipped protons to gain 63% of the net equilibrium magnetization is called T1 and the process is called T1 relaxation. The time taken for the excited protons in transverse plane to lose phase with each other and decay to 37% of the maximum transverse magnetization is called T2 and the process is called T2 relaxation. Additionally, the term T2* refers to T2 but it also takes into consideration the effects of local magnetic inhomogeneities in the field.[16] T2 relaxation time varies between tissues and the presence of iron has been linked with decreasing T2. If a normal tissue without iron experiences MRI signal decay in a particular amount of time, iron causes the MRI signals to decay even faster due to the magnetic field inhomogeneities introduced by the iron particles. Therefore, there is an inverse relation between time taken by the MRI signals to decay and iron concentrations; higher the iron concentration, shorter the T2. Alternately, iron concentration has a direct relationship with R2, which is the inverse of T2, and additionally with R2*, which is R2 combined with the effects of local field inhomogeneities. Relaxometry-based MRI techniques utilize this concept of relaxation to estimate the concentration of iron in different organs.[5]

Many studies on patients, phantoms, and simulations have been conducted to demonstrate the relationship between HIC and MRI-based T2, T2*, and R2*.[5-7] Calibration curves have been

proposed to standardize these relationships after performing MRI scans on patients, and calculating relaxation values, followed by liver biopsies on the same patients and determining HIC. These calibrations have been criticized for differing by the range of iron overload degree and the number of patients used. Primarily, these published calibrations have used a monoexponential signal model.[5,13] This means that they fit the MRI transverse signal decay to a monoexponential model and estimate R2*. Water signal follows a monoexponential fashion to dampen in the presence of iron only (Figure 2) and thus the monoexponential model can accurately estimate R2* in conditions of iron only.[5,6]

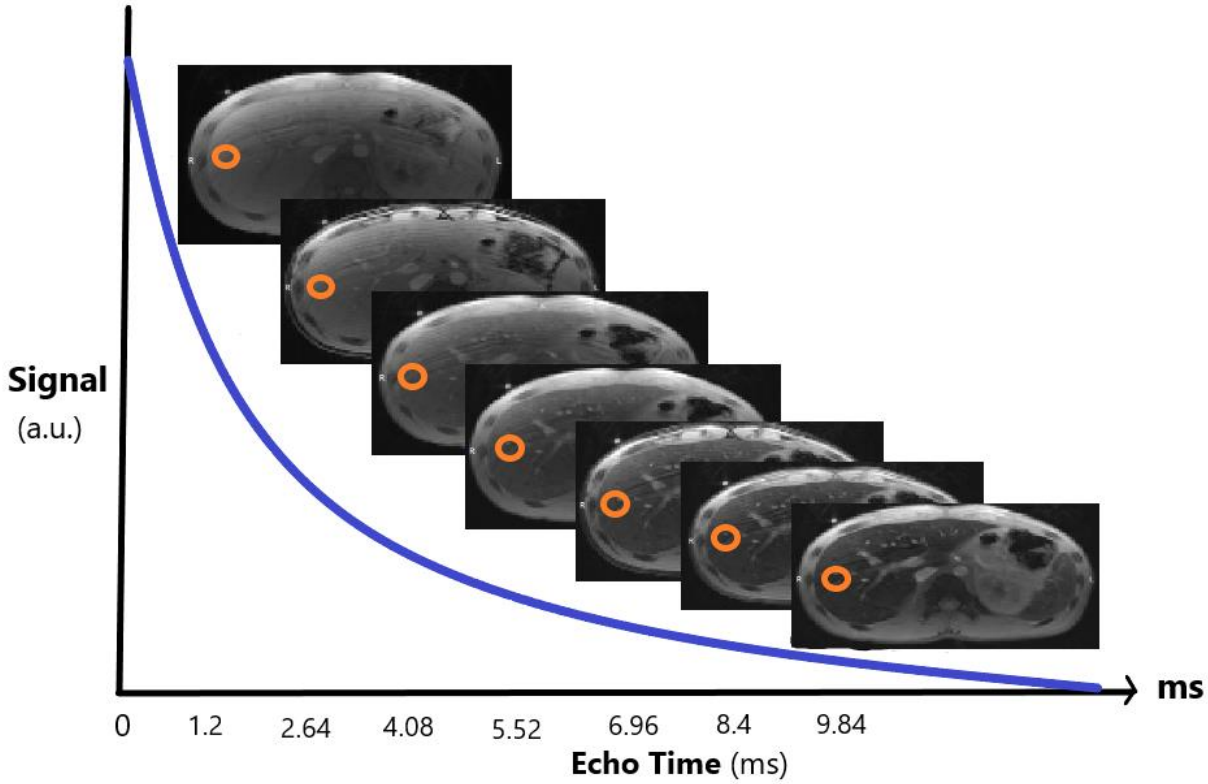

Figure 2. Representative example of a patient with iron overload scanned using MRI. At each echo time, the representative ROI is darker due to loss of signal intensity with increasing echo time (TE). Signals decays in a monoexponential fashion in the presence of iron only.

Hepatic fat accumulation, or steatosis, is a common condition, similar to iron overload. Water and fat are the key molecules in the human body. Like R2*, fat fraction (FF) is used as a biomarker to measure fat accumulation in the liver. Due to the complexity of triglycerides, their hydrogen protons experience a lesser shielding effect to the magnetic field. As a result, hydrogen protons in triglycerides resonate slightly slower than those in water , resulting in a frequency difference of around 220 Hz at 1.5 T and 440 Hz at 3T between the fat and water peaks.[8] When two signals have different frequencies, they oscillate in and out of phase, causing oscillations in the signal decay. When both iron and fat are present, the MRI signal shows oscillations that cannot be accurately described by a monoexponential model.[8] Since it's not always immediately known if a patient has both fat and iron overload, a model that accounts for the spectral complexity of fat and its effects on water signal is more robust and can be standardized for quantifying both R2* and FF.

Multispectral models such as NLSQ and ARMA have been proposed to account for the spectral complexity of fat and simultaneously quantify R2* and FF.[26,27] NLSQ is a non-linear least squares fitting technique that uses published information about the relative frequencies and amplitudes of multiple lipid peaks to fit a common R2* for both water and fat peaks.[26,28] Although fat and water peaks have similar R2* behavior, assuming they have the same R2* is technically inaccurate. Another signal model called the autoregressive moving average (ARMA) fixes this problem by converting the MRI signal to a rational polynomial in the z-domain to detect individual fat and water peaks and estimate independent R2* values for each signal.[8,27] There is a void in our understanding for which signal model accurately estimates R2* and FF and is the best approach for simultaneous quantification. In this study, we aim to investigate the performance of these two signal models. Since the monoexponential model is accurate for iron-

only conditions, it can serve as a reference model for comparing R2* estimation by NLSQ and ARMA.[5,6] We will also assess if the models can accurately detect the absence of fat peaks and estimate FF as 0% in our purely iron model.

Many MRI imaging sequences have been developed depending on the resolution of images, organs of interest, pathology of diseases, etc. Imaging sequences differ in their use of parameters like matrix size, field of view (FOV), storing of information in the k-space, echo times, etc. Echo time (TE), in simple terms, is the time taken by magnetic energy or radiofrequency (RF) pulse to return to the scanner after interacting with the body. The echo time refers to the time when MRI signals are gathered during every scan. The first signal is obtained at TE_1, and subsequently, signals are collected at each echo, with a certain interval between them (ΔTE), until reaching TEmax. TE values differs based on the type of tissue and pathology being visualized. TE can be considered as a form of signal sampling that affects the quality of MRI images. The conventional MRI acquisition for quantifying iron overload is the Cartesian-based 2D multiecho gradient echo (GRE) imaging sequence with initial echo time (TE_1) at approximately 1 ms.[11,29] However, this method may not accurately estimate R2* for moderate or massive iron overload cases when the signal decay is rapid and most of the signal might be lost before being captured by the initial echo of a GRE sequence. In contrast, the multiecho ultrashort echo (UTE) acquisition acquires data in a radial pattern, starting from the center of the k-space and moving outward. This method allows for a faster data acquisition and better motion correction. The use of a shorter TE_1 in UTE acquisition that can go as low as 0.1 ms has been shown to increase the dynamic range of R2* quantification in some studies.[10,11,29]

Previous studies have evaluated NLSQ and ARMA signal models in GRE acquisitions and found that both models accurately estimate R2* in mild and moderate iron overload

conditions.[8,26] However, in high iron overload conditions, these models have been shown to produce inaccurate results in some studies.[8,9] Additionally, the multispectral models till date have only been studied in GRE acquisitions, except for a recent study by Kee et. al that evaluated the NLSQ model in UTE acquisition at 3T UTE cones and found better performance compared to GRE.[12] However, this study was limited to a small cohort of patients with mild and moderate iron overload. Therefore, a comprehensive and systematic investigation of the multispectral models in different GRE and UTE acquisitions, with varying echo spacings (ΔTE) and maximum echo time (TEmax), is necessary to accurately assess the performance of NLSQ and ARMA models throughout the full clinically relevant range of HIC in iron overload patients.

Chapter 3 - Monte Carlo-based Simulation and Multispectral Fat-Water-R2* Models

Design of Virtual Iron Overload Model and MRI Signal Synthesis

A Monte Carlo simulation model is a computational method for generating a large number of random samples to simulate a real-world complex process or system. Random inputs are generated for one aspect of the model, and the model is run repeatedly getting results for a variety of different conditions and the results of each simulation run are then aggregated to estimate the output of the model as a whole. They have been used in a variety of applications like modeling of nanoparticles, organs composition, modeling mechanisms like entropy, etc. Monte Carlo simulations can create realistic virtual liver models by utilizing published statistics of hepatic iron scale, distribution, and magnetic properties and simulate MRI signals mimicking in vivo proton-iron interactions.[13] There is a void in our understanding of which fat-water-R2* model accurately estimates R2* and FF for the entire clinical range of iron overload. Further, since the MRI acquisitions used, especially in terms of echo times, play an important role in the performance of the signal models, it is also useful to see which acquisitions work the best for the signal models. Evaluation of fat-water models in multiple MRI acquisitions using patients or phantom models will be expensive and challenging. This gap can be bridged by simulation models that accurately illustrate the influence of varying iron content on MRI signals.

Monte Carlo model developed by Ghugre et. al for simulating virtual liver models in different iron overload conditions has gained credibility and been proven useful in the investigation of iron mediated relaxivity behavior.[13,31] Gamma Distribution Functions (GDFs) were used in this study to derive relations between HIC and size of iron deposits, nearest neighbor distance representing interparticle spacing between iron deposits (Figure 3), and cellular anisotropy representing different amount of iron deposits in different hepatocytes (Figure

4). Using these relations, GDFs for size, nearest neighbor, cellular anisotropy can be generated for any iron concentration. This Monte Carlo-based model was reproduced in our study to simulate virtual iron overload models mimicking realistic liver geometries. HICs ranging from 1 to 40 mg/g of dry liver weight were chosen to represent the clinical range of liver iron overload. The volume fraction of iron deposits was calculated as a function of HIC (Table 1) and the homogenous iron spheres were placed inside the liver volume. .[13,15]

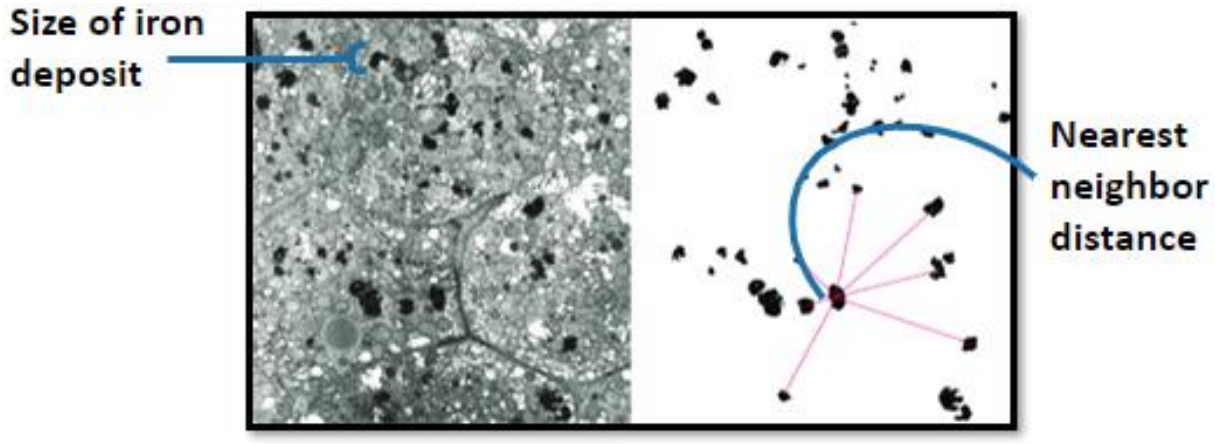

Figure 3. Electron micrograph for a representative patient with an iron burden of 16.6 mg/g dry tissue weight.[15] The image shows iron-filled lysosomes are indicated to demonstrate the size range of iron deposits used in simulations. The red lines extending from the iron deposit on the right indicate the distance between it and its neighbors, with the nearest neighbor (NN) being the shortest of these distances.

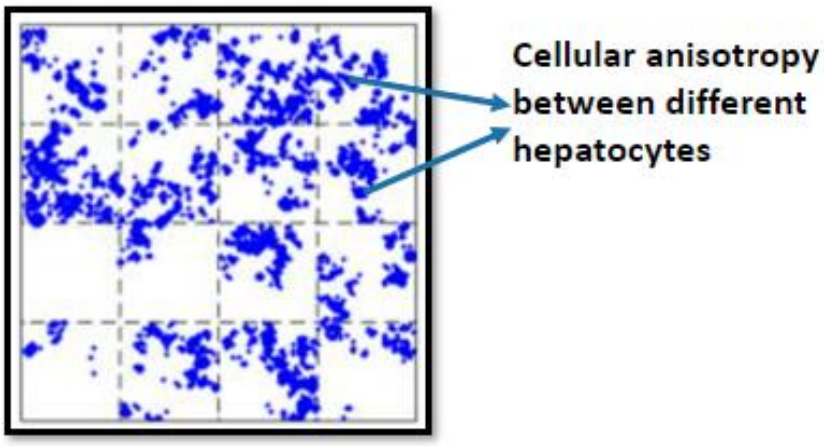

Figure 4. Representative 2D slice of dimension 80 µm * 80 µm of the virtual iron overload model, with indication of the difference in simulated distribution of iron deposition between each hepatocyte (cellular anisotropy).[13,15]

Table 1. Gamma distribution function (GDF) derived relations between HIC and volume fraction, size distribution, nearest neighbor, and cellular anisotropy of iron deposits.[13]

Characterization	Equation	m [units]	c [units]	RMSE	R
Volume fraction	$VF = m^*HIC+c$	0.00127 $[(mg/g)^{-1}]$	0.00368 [%]	0.00716	0.93
Size distribution	$\beta = m^*HIC+c$	0.000296 $[\mu m.(mg/g)^{-1}]$	0.0935 [µm]	0.0245	0.17
	$\gamma = m^*\beta+c$	2.188 $[\mu m^{-1}]$	2.695 [ ]	0.0126	0.97
Nearest neighbor	$\gamma = \exp(m^*\log(HIC)+c)$	0.476 $[(mg/g)^{-1}]$	−0.131 [ ]	1.221	0.78
	$\beta = \exp(m^*\log(\gamma)+c)$	−1.403 [µm]	0.740 [µm]	0.357	0.98
Cellular anisotropy	$\gamma = m^*HIC+c$	0.0179 $[(mg/g)^{-1}]$	1.025 [ ]	0.412	0.53
	$\beta = \exp(m^*\log(HIC)+c)$	0.375 $[(mg/g)^{-1}]$	−3.366 [ ]	0.0296	0.76

Abbreviations: β, Scale parameter; γ, Shape parameter; RMSE, Root Mean Square Error; R, Correlation coefficient; VF, Volume Fraction of spheres (%); HIC, Hepatic Iron Concentration (mg/g dry tissue weight); [ ]: Dimensionless

MRI signals were simulated by generating an inhomogeneous magnetic field induced by the iron deposits and allowing 5000 protons to randomly move across the magnetic field. Superposition of phase accrual of all protons as they pass through the induced magnetic field gave the corresponding MRI signal. Susceptibility value of 1.6 E-6 m^3/kgFe was set equivalent to the 4:1 mixture of human hemosiderin and ferritin.[13] Expected magnetic field inhomogeneities induced by the iron deposits were modeled using:

$$\Delta B(r,\theta) = \left(\frac{B0}{3}\right)\chi\left(\frac{R}{r}\right)^3 (3Cos^2\theta - 1)$$

where B0 is the applied magnetic field, χ is the particle susceptibility, R is the sphere radius, r is the radial distance between the center of the iron sphere and proton position, and θ is the azimuthal angle to the magnetic axis. Proton mobility was modeled in terms of mean displacement given by:

$$\sigma = \sqrt{2D\delta}$$

where D=0.76 μm^2/ms is the diffusion coefficient and δ=0.5 μs is the proton time step.[32] The phase accrual ϕ for each proton at the end of each time step t was calculated using:

$$\emptyset(t) = \gamma\delta\sum_{i=1}^{t}\left(B_0 + \triangle B\left(p(i)\right)\right)$$

where γ=2.675*108 rad s-1 T-1 is the gyromagnetic ratio and p(i) is the i^{th} proton position. Finally, the complex MRI signal from each proton is given by:

$$S(t) = S(0)e^{-t*R_{2,0}+j\phi(t)}$$

where S(0) is signal amplitude at t=0 and $R_{2,0}$ is the relaxation in normal liver, assumed to be 20 s^{-1} at 1.5 T and 35 s^{-1} at 3 T.

Signal Models for R2* Estimation

After the MRI signals were synthesized by varying echo times (TEmin, ΔTE, TEmax) for different acquisitions, they were fitted to monoexponential and multispectral models to obtain relaxivity values. Monoexponential model, implemented from Wood et. al. study, uses the relation $S(t) = S0 * e^{\left(-\frac{t}{T_2*}\right)} + c$ to estimate R2*,[5] where S(0) is the initial signal, S(t) is the complex MRI signal at that time, and R2* is the inverse of T2*. ARMA and NLSQ models, on the other hand, are more complex.[8,27,28] The NLSQ model leverages prior information by utilizing published values for the relative frequencies and amplitudes of the lipid peaks.[28] This model fits a single R2* value for both water and fat peaks to reduce model complexity.[26,28] By using the following equation that represents mGRE signal acquired in a single voxel at each echo time (TE_n), incorporating the provided parameters for amplitudes and relative frequencies, and assuming that the R2* values for water and lipid peaks are equal, the NLSQ model is able to estimate a common or single R2*.[8]

$$S(TE_n) = \left(C_W e^{-(i2\pi f_W + R^*_{2,W})TE_n} + C_F \sum_{m=1}^{M} \alpha_m e^{-(i2\pi f_{F,m} + R^*_{2,F,m})TE_n} \right)$$

where, C_W and C_F are the amplitudes of the water and lipid signals, respectively. Additionally, $R^*_{2,W}$ and $R^*_{2,F,m}$ denote the R2* values for water and lipid peaks, respectively. The variable *M* represents the number of lipid peaks, while f_W and $f_{F,m}$ represent the relative frequencies for water and lipid peaks, respectively. Finally, α_m denotes the relative amplitudes of multiple lipid peaks, such that the sum of all α_m equals 1.

On the other hand, the ARMA model estimates independent or dual R2* for water and fat peaks. It uses no prior information and estimates amplitude, R2*, and field maps of each detected peak by converting the MRI signal to a rational polynomial in the z-domain.[8,27] The above equation of mGRE signal can be rewritten in the form of a autoregressive moving average by using a comb function as follows:[8]

$$S(t) = \left(C_W e^{-(i2\pi f_W + R^*_{2,W})t} + C_F e^{-(i2\pi f_F + R^*_{2,F})t}\right) * \sum_{t_i = TE_1}^{TE_1 + (n-1)\Delta TE} \delta(t - t_i)$$

ARMA model then applies a z-transform to the mGRE signal, which can be rewritten as follows:

$$S(z) = \sum_{t_i = TE_1}^{TE_1 + (n-1)\Delta TE} C_W e^{-(i2\pi f_W + R^*_{2,W})t} z^{-t} + \sum_{t_i = TE_1}^{TE_1 + (n-1)\Delta TE} C_F e^{-(i2\pi f_F + R^*_{2,F})t} z^{-t}$$

Finite summation with the sum of the geometric series is approximated.[8]

$$S(z) = \frac{C_W}{1 - e^{-(i2\pi f_W + R^*_{2,W})\Delta TE} z^{-1}} + \frac{C_F}{1 - e^{-(i2\pi f_F + R^*_{2,F})\Delta TE} z^{-1}}$$

This equation can be extended to *N* number of peaks and the mGRE signal can be represented in the z-domain as:

$$S(z) = \frac{P(z)}{Q(z)} = \frac{\alpha_0 + \alpha_1 z^{-1} + \cdots + \alpha_{N-1} z^{-(N-1)}}{1 + \beta_1 z^{-1} + \cdots + \beta_N z^{-N}}$$

The iterative Stieglitz-McBride algorithm can calculate the coefficients α and β for the two polynomials *P(z)* and *Q(z)*.[8] Following this calculation, the poles are then derived from the roots of the denominator *Q(z)*. Subsequently, the frequencies, *R2** values, and amplitudes of the water and fat peaks can be extracted from the poles and coefficients using the following approach:

$$f_k = \frac{Im[\ln(\rho_k)]}{2\pi.\Delta TE.\gamma B_0}$$

$$R^*_{2,k} = \frac{Re[\ln(\rho_k)]}{\Delta TE}$$

$$C_k = \frac{P(\rho_k)}{Q'(\rho_k)} \; where \; Q'(z^{-1}) = \frac{dQ(z^{-1})}{dz^{-1}}$$

where, *Re* and *Im* denote real and imaginary parts, respectively. γ is the gyromagnetic ratio and B_0 is the main magnetic field strength.[8]

In our study, NLSQ model was implemented from the ISMRM Fat-Water Toolbox and ARMA model was implemented as an iterative approach with a maximum of 7 peaks.[8,28]

Chapter 4 - Simulation of a Virtual Liver Iron-Overload Model and R2* Estimation Using Multispectral Fat-Water Models for GRE and UTE Acquisitions

Introduction

Iron overload is a serious condition caused by hereditary hemochromatosis or due to receiving chronic blood transfusions for the management of hematologic diseases such as sickle cell disease, beta-thalassemia or as part of cancer treatment.[1,2] Iron overload can cause hepatic fibrosis and cirrhosis, hepatocellular carcinoma, cardiac problems, and many other organ dysfunctions if untreated.[1,33] Liver iron stores increase proportionally with total body iron, so assessing hepatic iron concentration (HIC) is an effective way of monitoring iron overload for treatment management.[4] Traditionally, HIC is assessed using liver biopsy which is an invasive, painful, and risky procedure.[5] Alternatively, over the last decade, magnetic resonance imaging (MRI) techniques based on transverse relaxation (R2*) have emerged as clinically accepted methods for quantifying HIC using published R2* vs. biopsy HIC calibration curves.[4-7] Previous in vivo calibration studies estimated R2* by acquiring multiecho gradient echo (GRE) images and fitting a monoexponential signal model.[5,6] However, pure monoexponential models can cause inaccuracies in R2* estimation in the presence of concomitant hepatic steatosis as fat introduces oscillations in the signal decay.[8]

Over recent years, multispectral fat-water-R2* techniques based on non-linear square (NLSQ) fitting and autoregressive moving average (ARMA) modeling have been proposed for simultaneously quantifying R2* and fat fraction (FF) to account for both the confounding effects of iron and fat on the MRI signal.[8,27,28] The NLSQ signal model assumes relative frequencies and amplitudes of the lipid peaks and fits a common R2* value to both the water and fat peaks to minimize model complexity,[26,28] and has been validated for simultaneous quantification of iron and fat in simulations, phantoms and in vivo studies.[9,33,34] The ARMA signal model, on the other

hand, detects signals based on the frequencies of water and lipid peaks and calculates independent R2* values and amplitudes for each detected peak.[27] A previous study in phantoms and patients has shown that both ARMA and NLSQ models produced inaccurate R2* results in some conditions of high iron overload (R2* > 500 s^{-1}) with and without steatosis, respectively; both methods potentially failing due to rapid R2* decay that is not adequately captured with the current GRE acquisitions with TE_1 of ~1 ms.[8]

Alternatively, multiecho ultrashort echo time (UTE) acquisitions with shorter TE_1 (0.1-0.19 ms) have been proposed for quantifying severe iron overload and demonstrated to improve the accuracy and precision in R2* quantification over a wider range of clinical HIC values at both 1.5 T and 3 T using monoexponential models.[10,11,29] To date, multispectral fat-water models have not been thoroughly investigated for estimating R2* for UTE acquisitions, except one study that incorporated NLSQ multispectral model for R2* quantification using a 3D UTE cones acquisition at 3 T.[12] However, this study had only patients with mild to moderate iron overload (R2* < 550 s^{-1}), and the NLSQ R2* results were not validated with any reference monoexponential fitting model or biopsy HIC values. A thorough and systematic investigation of the multispectral fat-water-R2* models for different GRE and UTE acquisitions will be beneficial to assess their performances under various hepatic iron overload conditions before clinical use. But evaluating fitting models across multiple MRI acquisitions in phantoms or patients can be expensive and challenging. Instead, some previous studies implemented Monte Carlo simulations, based on published in vivo statistics of hepatic iron scale, distribution, and magnetic properties, and have been successful in simulating realistic virtual liver iron overload models and generating synthetic MRI signals mimicking in vivo behavior.[13-15,31]

The purpose of our study is therefore, to construct a virtual liver model with varying hepatic iron concentrations using Monte Carlo simulations to evaluate the performances of multispectral fat-water signal models for different GRE and UTE acquisitions and validate their accuracies against reference monoexponential model and published in vivo R2*-HIC calibrations.

Methods

Monte Carlo model developed by Ghugre et. al was implemented in our study for simulation of a virtual iron overload model and synthesis of MRI signals.[13-15] An 80 μm * 80 μm * 80 μm liver volume with 64 hepatocytes, each with 20 μm dimensions, was simulated. Based on previously derived relationships, volume fraction of the iron deposits was calculated as a linear function of the desired HICs ranging from 1 - 40 milligrams Fe per gram of dry weight of liver (abbreviated as mg Fe/g). Iron deposits were modeled as homogeneous spheres, with their sizes, nearest neighbor distances, and cellular anisotropy across hepatocytes obtained from respective gamma distribution functions derived by Ghugre et al. as a function of HIC. The iron spheres were placed within hepatocytes in the cuboidal liver volume to create virtual iron overload models with different HICs until the corresponding desired volume fraction was reached for each HIC.[13] Iterative collision check was performed to ensure that no iron deposits overlapped.[13]

MRI signals were synthesized at both 1.5 T and 3 T by calculating the magnetic field inhomogeneities induced by iron deposits, simulating the proton mobility, computing the phase accrual of protons, and finally converting the phase differences into MRI signal. A total of 5000 protons were randomly distributed in the liver volume and they all performed a random walk following unrestricted diffusion for a total duration of 10 ms. For each HIC, the synthetic complex MRI signal was produced by superimposing the phase accrual of all protons as they passed through the iron-induced magnetic field disturbances. Monte Carlo simulations for creating virtual liver

iron overload models and generating MRI signals at 1.5 T and 3 T for HICs ranging from 1 - 40 mg Fe/g were all performed using an in-house script developed in Python (Python 3.8.8, Python Software Foundation, Wilmington, DE, USA.).

Data Analysis

MRI signals were synthesized for GRE and UTE acquisitions for varying sets of echo spacing, ΔTE (0.1, 0.5, 1, 2 ms) and maximum echo time, TEmax (2, 4, 6, 10 ms) for all iron overload models with HICs 1 – 40 mg Fe/g, incremented by 1 mg Fe/g. R2* values were calculated using monoexponential and multispectral fat-water signal models in MATLAB (R2020b, The MathWorks, Inc., Natick, MA, USA). A monoexponential model with constant offset was used to fit the magnitude data and estimate R2* values.[5,13] An iterative ARMA model with a maximum of 7 peaks and the NLSQ model from ISMRM Fat-Water Toolbox were applied on the complex MRI signal data to estimate R2* and FF values.[8,28] In vivo R2*-HIC calibration curves published by Wood et al. and Hankins et al. were digitized and used as reference standards to compare the accuracies of our model estimated R2* vs. HIC relationships for comparable GRE acquisition parameters (ΔTE = 0.5 ms, TEmax = 10 ms).[5,6,13]

Statistical Analysis

For each signal model, R2* values estimated for different GRE and UTE acquisitions with varying sets of ΔTE and TEmax were analyzed across the entire range of simulated HICs. Simulation for each acquisition setting was repeated three times and the mean R2* and FF was taken in order to reduce bias introduced by random variations in the simulations. The precision of the fitting models was evaluated by using the coefficient of variation (CoV, %) of R2* values across the three simulations plotted against HICs. The accuracy of the signal models was evaluated by using linear regression analysis between the mean estimated R2* values and simulated HICs for the various GRE and UTE acquisitions.

Results

Representative 3D virtual iron overload models created based on the published iron morphology and distribution for mild and moderate HICs and the corresponding MRI signals synthesized at 1.5 T and 3 T are shown in Figure 5. The simulated MRI signals exhibited exponential signal decay with faster decay rates for higher HIC and at higher field strength. Figure 6 shows R2* values obtained using the monoexponential, ARMA, and NLSQ models for GRE acquisition at 1.5 T and comparison to in vivo R2*-HIC calibrations. The R2* values obtained using the monoexponential model exhibited an excellent correlation with HIC values, with slopes within 95% confidence interval of Wood et al. calibration. Both ARMA and NLSQ model estimated R2* values also exhibited an excellent correlation with HIC values, with slopes higher than monoexponential model and published R2*-HIC calibrations but within 95% confidence interval of Wood's calibration except for the slope of the NLSQ model for higher HICs (> 20 mg Fe/g).

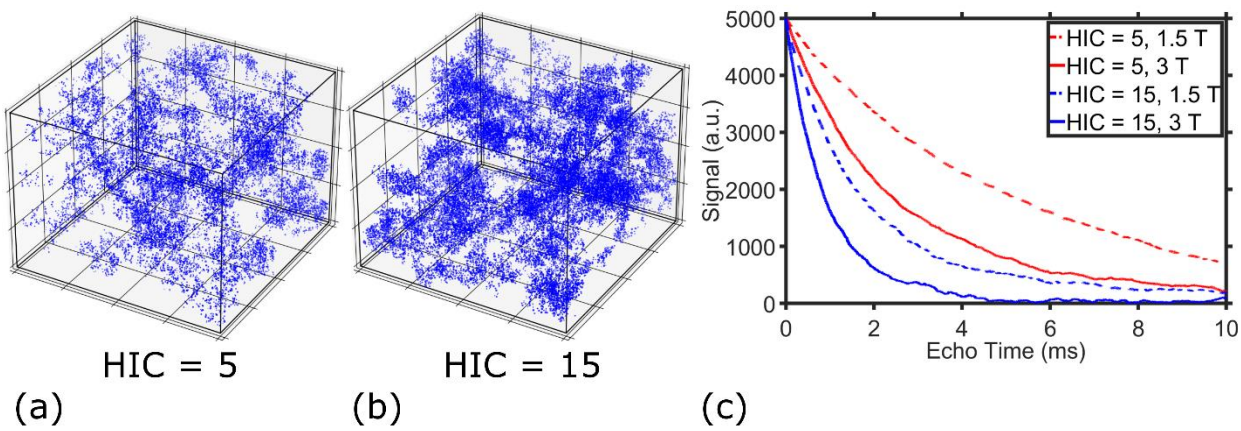

Figure 5. 3D virtual liver iron overload models (80 µm * 80 µm * 80 µm) for representative HICs of 5 mg Fe/g (a), 15 mg Fe/g (b), and corresponding MRI signals at 1.5 T and 3 T for a duration of 10 ms (c). The signals showed faster signal decay at higher HIC (R2* = 220 s^{-1} for 5 mg Fe/g and R2* = 591 s^{-1} for 15 mg Fe/g at 1.5 T) and field strength (R2* = 433 s^{-1} for 5 mg Fe/g and R2* = 1032 s^{-1} for 15 mg Fe/g at 3 T), as expected.

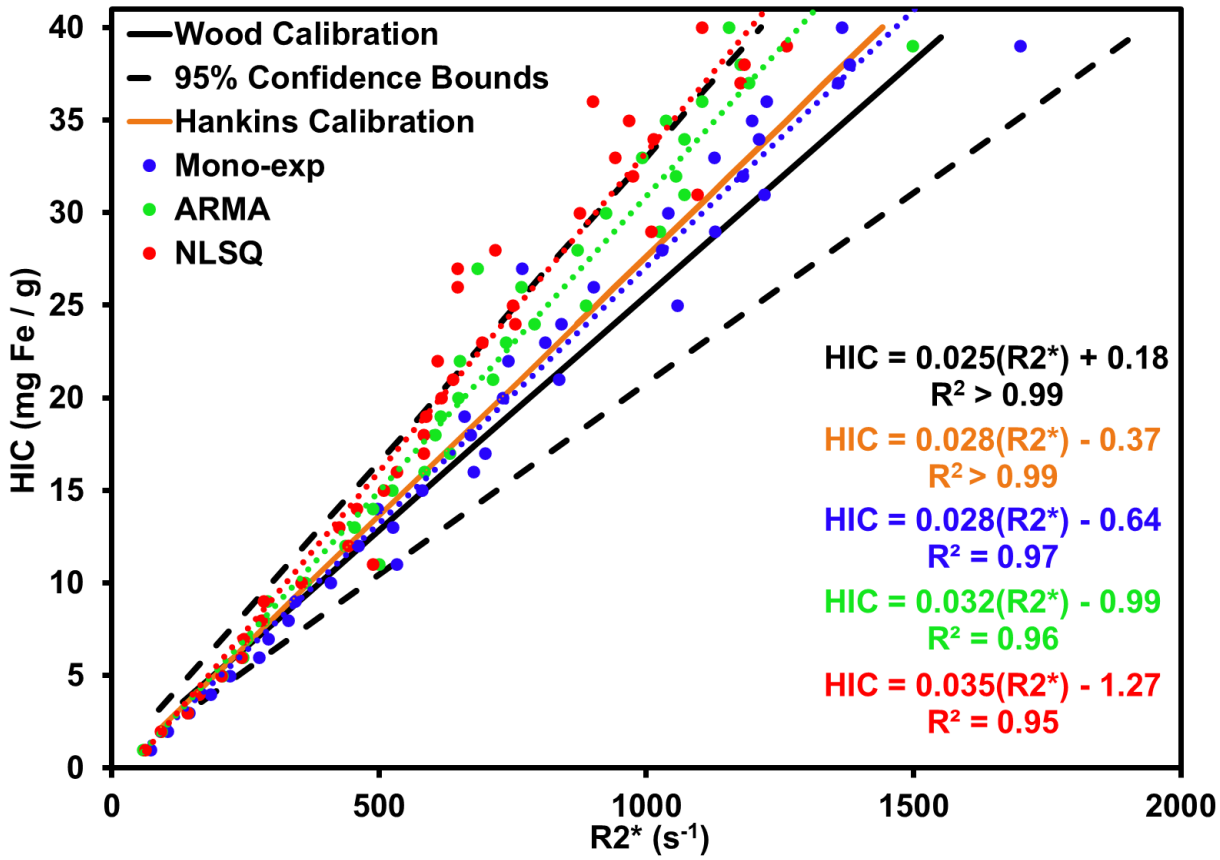

Figure 6. Linear regression between R2* values estimated using monoexponential, ARMA, and NLSQ models and simulated HIC values, in comparison to in vivo R2*-HIC calibrations for GRE acquisition at 1.5 T. R2* values estimated using the monoexponential model exhibited excellent correlation with HIC values; the R2*-HIC slope is same as Hankins et al. calibration and falls within 95% confidence bounds of Wood et al. calibration (dashed lines). ARMA and NLSQ model estimated R2* values also showed excellent correlation with HICs, with slopes higher than the monoexponential model and in vivo calibrations but within Wood et al. 95% confidence limits, with the exception of the NLSQ model at higher HICs (> 20 mg Fe/g).

The R2* values estimated using monoexponential, ARMA, and NLSQ models for GRE and UTE acquisitions for various ΔTEs and constant TEmax of 10 ms at both 1.5 T and 3 T for the simulated HIC values are shown in Figure 7 and the corresponding regression slopes are listed in Table 2. At 1.5 T, GRE acquisition produced similar R2* results for all ΔTEs for each of the 3 signal models up to HIC < 15 mg Fe/g; however, deviation from the linear trend was observed with the increase in ΔTEs for higher HICs for all 3 signal models with the NLSQ model showing the greatest deviation at ΔTE = 2 ms. In comparison to GRE, the estimated R2* values by monoexponential and ARMA models for UTE acquisition at 1.5 T demonstrated relatively low sensitivity to variations in ΔTE for the entire HIC range (Table 2, Slope: 0.028-0.030, R^2: 0.982-0.991). NLSQ model also demonstrated to be less sensitive to variations in ΔTEs up to HIC ~20 mg Fe/g for UTE, but greater deviations in R2* estimates were observed for ΔTE = 2 ms for higher HICs. At 3 T, the estimated R2* values by monoexponential and ARMA models for GRE acquisition demonstrated excellent linear correlation for lower HICs (≤ 12.5 mg Fe/g) at all ΔTEs but for higher HICs, the R2* values did not increase linearly with HIC for all ΔTEs (Table 2, R^2: 0.356-0.789). Similarly, NLSQ estimated R2* values for GRE acquisition at 3 T exhibited higher levels of deviations for all ΔTEs but from even lower HICs (≥ 10 mg Fe/g). In contrast, UTE at 3 T produced similar R2* results for all ΔTEs using monoexponential and ARMA signal models, except that at ΔTE = 2 ms, estimated R2* values showed greater deviations at higher HICs (> 30 mg Fe/g). NLSQ model estimated R2* values also exhibited excellent linear correlation for smaller ΔTEs (≤ 0.5 ms), however, for larger ΔTEs ≥ 1 ms, the R2* values deviated from linear trend, with ΔTE = 2 ms showing extreme divergence for HICs > 10 mg Fe/g (Table 1, R^2 = 0.723).

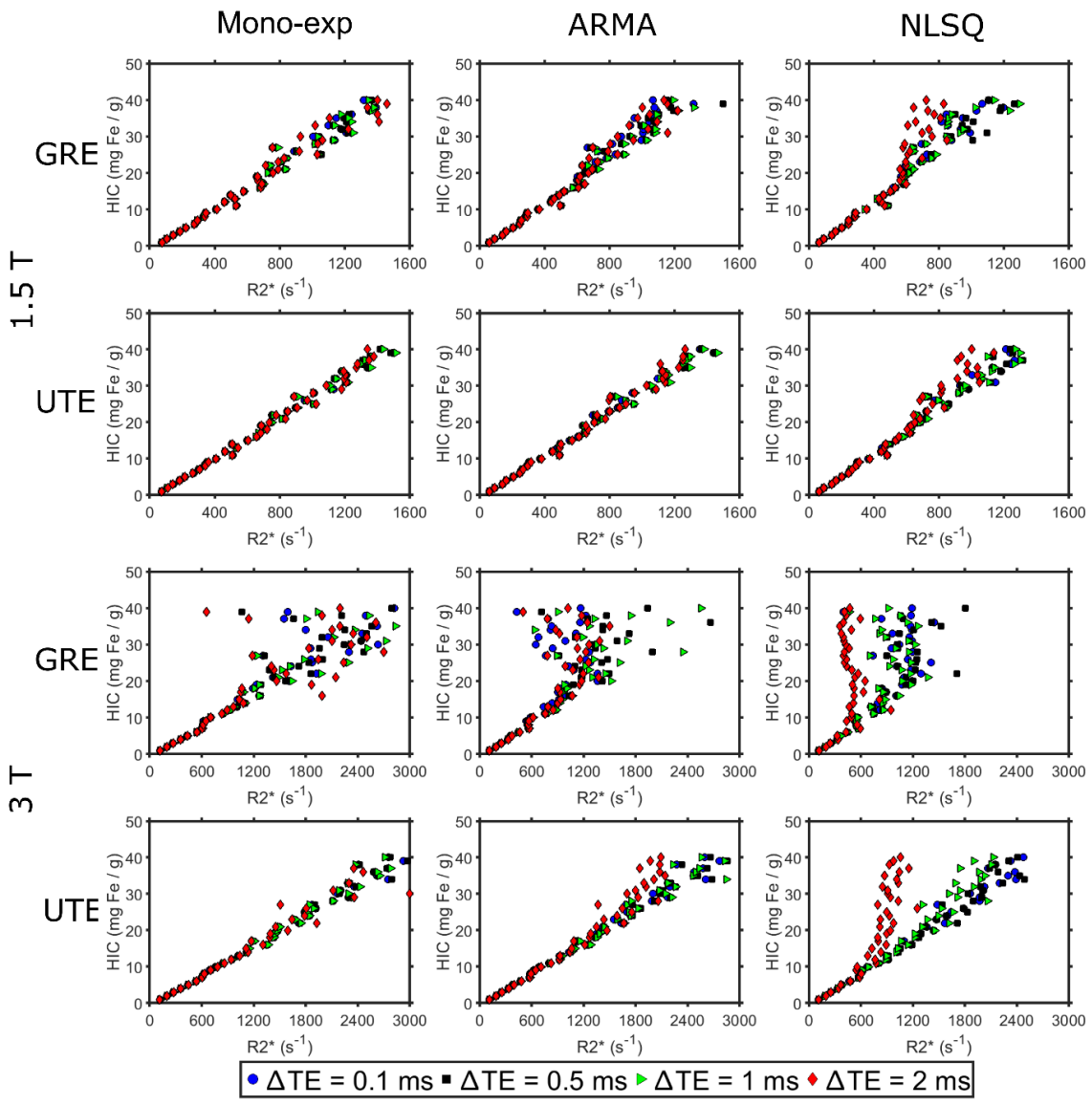

Figure 7. Scatter plot demonstrating the effect of varying ΔTE (0.1, 0.5, 1, 2 ms) with constant TEmax (10 ms) on R2*-HIC relationship obtained using monoexponential, ARMA, and NLSQ signal models for GRE and UTE acquisitions at 1.5 T and 3 T. UTE acquisition with $\Delta TE \leq 0.5$ ms produced accurate R2* results as well as similar R2*-HIC relationships (Table 2) for all three signal models across the full range of simulated HIC at both 1.5 T and 3 T.

Table 2. Linear regression analysis between R2* values estimated by different signal models and simulated HICs for GRE and UTE acquisitions for various echo spacings at 1.5 T and 3 T.

Signal Model	Parameters	1.5 T								3 T							
		GRE				UTE				GRE				UTE			
		ΔTE (ms)				ΔTE (ms)				ΔTE (ms)				ΔTE (ms)			
		0.1	0.5	1	2	0.1	0.5	1	2	0.1	0.5	1	2	0.1	0.5	1	2
Mono-exp	Slope	0.028	0.028	0.024	0.024	0.028	0.028	0.028	0.028	0.012	0.011	0.009	0.012	0.014	0.014	0.014	0.008
	Intercept	-0.572	-0.645	1.538	2.009	-1.260	-1.381	-1.246	-1.189	2.786	-4.013	-5.741	3.702	-0.990	-0.942	-0.448	-5.606
	R^2	0.958	0.965	0.897	0.865	0.990	0.991	0.991	0.985	0.789	0.726	0.737	0.618	0.986	0.986	0.971	0.778
ARMA	Slope	0.035	0.032	0.029	0.035	0.029	0.029	0.029	0.030	0.021	0.018	0.014	0.023	0.015	0.015	0.015	0.018
	Intercept	-2.075	-0.988	0.392	-2.905	-0.807	-0.841	-0.772	-0.991	3.154	-1.761	-5.319	0.057	-0.787	-0.676	-0.716	-3.156
	R^2	0.964	0.963	0.934	0.956	0.988	0.990	0.990	0.982	0.356	0.660	0.567	0.480	0.985	0.986	0.973	0.945
NLSQ	Slope	0.038	0.035	0.036	0.051	0.032	0.032	0.032	0.036	0.243	0.022	0.024	0.005	0.016	0.016	0.019	0.038
	Intercept	-2.434	-1.269	-1.794	-5.946	-1.246	-1.295	-1.212	-2.138	-0.537	-0.553	-0.956	-18.144	-1.608	-1.642	-3.052	-9.353
	R^2	0.954	0.952	0.942	0.853	0.977	0.981	0.981	0.927	0.532	0.557	0.414	0.004	0.978	0.969	0.942	0.723

Abbreviations: Mono-exp, monoexponential model; ARMA, autoregressive moving average model; NLSQ, non-linear least squares model; GRE, gradient echo; UTE, ultra-short echo; ΔTE, echo spacing.

The CoV of estimated R2* values obtained using monoexponential, ARMA, and NLSQ models across different HICs and ΔTEs for GRE and UTE acquisitions (Figure 8) showed similar trends as the estimated mean R2* values in Figure 3. At 1.5 T, monoexponential and ARMA models for GRE acquisitions exhibited CoV < 20% for HICs < 20 mg Fe/g. The precision decreased at higher HICs (> 20 mg Fe/g at 1.5 T and > 10 mg Fe/g at 3 T) for all models for GRE acquisition, with NLSQ model showing relatively lower R2* precision than monoexponential and ARMA models. For UTE at 1.5 T, both monoexponential and ARMA models showed high precision (CoV < 15%) in R2* values throughout the range of HICs except for a few data points at ΔTE = 2 ms and higher HICs. NLSQ demonstrated similar high precision trends, except that the overall CoV of R2* values at all ΔTEs were higher than in the other two models, particularly for higher HICs (> 20 mg Fe/g). Similarly, at 3 T, all 3 signal models for UTE acquisition exhibited high precision (CoV < 20%) in R2* values when HIC < 20 mg Fe/g, except for NLSQ model at larger ΔTE of 2 ms. While precision decreased at higher HICs with increasing ΔTE, overall CoV remained less than 30% for all three models, except for a few data points at ΔTE of 2 ms.

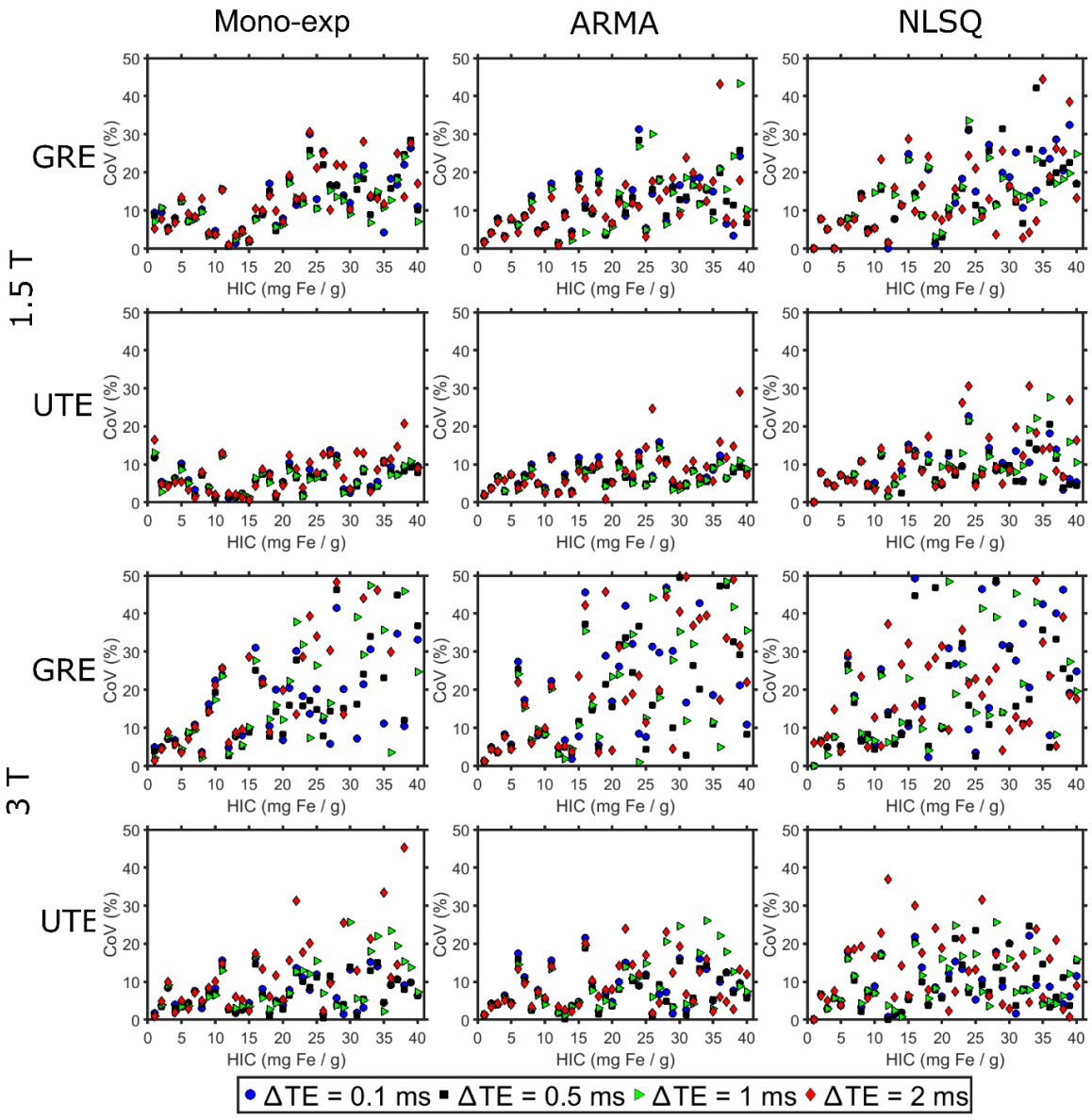

Figure 8. Coefficient of variation (CoV, %) of R2* values obtained using monoexponential, ARMA, and NLSQ models plotted against HIC values for GRE and UTE acquisitions with varying ΔTE (0.1, 0.5, 1, 2 ms) and constant TEmax (10 ms) at field strengths 1.5 T and 3 T. All signal models produced high R2* precision (i.e., lower CoVs) for UTE acquisition with ΔTE ≤ 0.5 ms across the full HIC range at both 1.5 T and 3 T, with ARMA and monoexponential models exhibiting slightly lower CoV values compared to NLSQ model.

Figure 9 shows R2* values obtained using monoexponential, ARMA, and NLSQ signal models and simulated HIC values for GRE and UTE acquisitions for various TEmax values at 1.5 T and 3 T and Table 3 shows the corresponding linear regression results. For GRE acquisitions at both 1.5 T and 3 T, decreasing TEmax from 10 ms did not affect R2* vs. HIC regression trends until TEmax of 4 ms for all signal models. However, at TEmax = 2 ms, monoexponential model produced greater deviations in R2* values across the entire HIC range, whereas ARMA and NLSQ models exhibited greater R2* deviations at HIC > 20 mg Fe/g, with the NLSQ model producing higher deviations than the ARMA model. UTE at 1.5 T and 3 T, on the other hand, still demonstrated excellent linear correlation between estimated R2* values and simulated HIC for all signal models and at all TEmax values across the full HIC range.

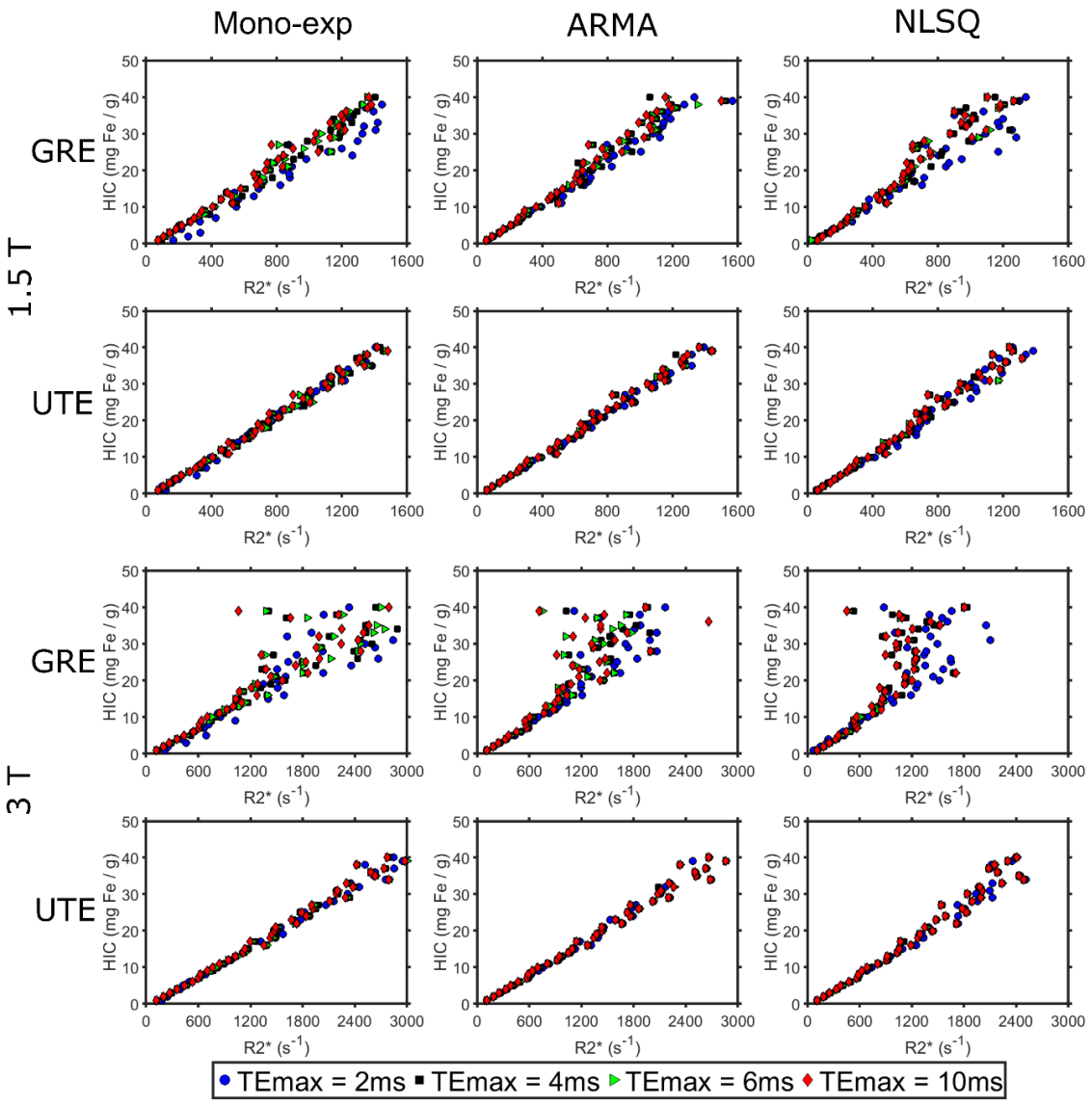

Figure 9. Scatter plot demonstrating the effect of varying TEmax (2, 4, 6,10 ms) with constant echo spacing (0.5 ms) on R2*-HIC relationship obtained using monoexponential, ARMA, and NLSQ signal models for GRE and UTE acquisitions at 1.5 T and 3 T. For GRE acquisitions, decreasing TEmax did not affect the R2*-HIC trends for all signal models until 4 ms but for TEmax = 2 ms all models showed considerable R2* deviations. In contrast, all signal models produced accurate R2* results for UTE acquisition for all TEmax values at both 1.5 T and 3 T.

Table 3. Linear regression analysis between R2* values estimated by different signal models and simulated HICs for GRE and UTE acquisitions for various TEmax values at 1.5 T and 3 T.

Signal Model	Parameters	1.5 T								3 T							
		GRE				UTE				GRE				UTE			
		TEmax (ms)				TEmax (ms)				TEmax (ms)				TEmax (ms)			
		2	4	6	10	2	4	6	10	2	4	6	10	2	4	6	10
Mono-exp	Slope	0.025	0.027	0.028	0.028	0.029	0.028	0.028	0.028	0.008	0.011	0.011	0.011	0.014	0.014	0.014	0.014
	Intercept	-1.782	1.265	1.005	0.645	2.120	1.746	1.553	1.381	5.837	3.912	4.583	4.013	1.425	1.285	1.153	0.942
	R^2	0.946	0.968	0.965	0.965	0.992	0.991	0.992	0.991	0.704	0.724	0.723	0.726	0.989	0.985	0.986	0.986
ARMA	Slope	0.029	0.031	0.031	0.032	0.029	0.030	0.029	0.029	0.015	0.017	0.016	0.018	0.015	0.015	0.015	0.015
	Intercept	-1.225	1.228	0.783	0.988	1.331	1.212	0.976	0.841	2.115	1.252	3.086	1.761	1.378	0.925	0.764	0.676
	R^2	0.975	0.952	0.964	0.963	0.993	0.989	0.991	0.990	0.720	0.734	0.641	0.660	0.983	0.985	0.986	0.986
NLSQ	Slope	0.027	0.033	0.034	0.035	0.031	0.031	0.032	0.032	0.019	0.023	0.022	0.022	0.016	0.016	0.016	0.016
	Intercept	0.244	1.262	1.292	1.269	1.945	1.444	1.335	1.295	0.105	0.426	0.236	0.553	1.944	1.928	1.763	1.642
	R^2	0.916	0.937	0.949	0.952	0.983	0.978	0.978	0.981	0.680	0.569	0.558	0.557	0.966	0.969	0.969	0.969

Abbreviations: Mono-exp, monoexponential model; ARMA, autoregressive moving average model; NLSQ, non-linear least squares model; GRE, gradient echo; UTE, ultra-short echo; TEmax, maximum echo time.

The trends in the R2* precision obtained from the CoV analysis of different TEmax values for GRE and UTE acquisitions (Figure 10) were comparable to the estimated mean R2* values in Figure 9. For GRE acquisitions at both 1.5 T and 3 T, using shorter TEmax values (2 ms, 4 ms) reduced R2* precision for monoexponential and NLSQ models for some HICs, but notably more for lower HICs (< 10 mg Fe/g) for monoexponential model. In contrast, using shorter TEmax in ARMA model still produced similar CoV results throughout the entire HIC range. For UTE acquisitions at 1.5 T, decreasing TEmax produced similar CoV results in all three models throughout the full HIC range, with the exception of monoexponential and NLSQ models at TEmax of 2 ms and 4 ms at lower HICs (< ~10 mg Fe/g). In contrast, for UTE at 3 T, precision of monoexponential and NLSQ models for shorter TEmax (2 ms, 4 ms) improved for smaller HICs compared to 1.5 T, and all the signal models exhibited high precision (CoV < 25%) throughout the HIC range.

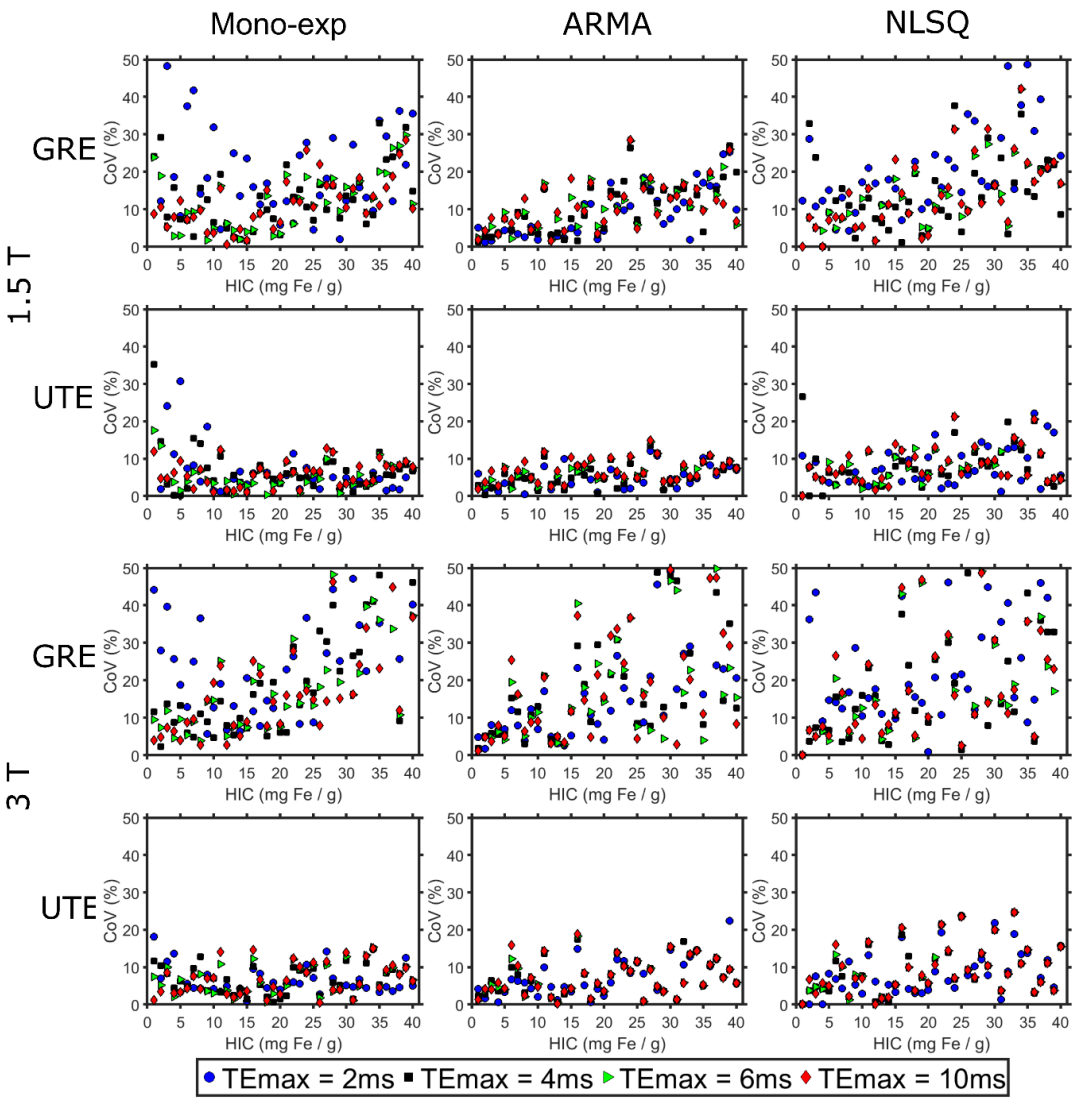

Figure 10. Coefficient of variation (CoV, %) of R2* values obtained using monoexponential, ARMA, and NLSQ models plotted against HIC for GRE and UTE acquisitions with varying TEmax and constant ΔTE at field strengths 1.5 T and 3 T. For GRE acquisitions, shorter TEmax ≤ 4 ms produced higher CoV values especially evident for monoexponential and NLSQ models. For UTE acquisition, all three signal models produced high precision in R2* values at both 1.5 T and 3 T across the full spectrum of HIC range for TEmax ≥ 6 ms, with ARMA model exhibiting similar precision trends as the monoexponential model.

Figure 11 displays the mean FF (%) values calculated using the NLSQ and ARMA models plotted against simulated HICs at 1.5 T and 3 T using one representative GRE and UTE acquisition (ΔTE = 0.5 ms, TEmax = 10 ms). At both 1.5 T and 3 T, the NLSQ model failed to estimate accurate FF, especially for HICs > 10 mg Fe/g, for both GRE and UTE acquisitions. Estimated FF values by the NLSQ model ranged from -12 – 168 % with a mean bias of 30 %. ARMA, on the other hand, estimated FF values close to 0% throughout the HIC range for both GRE and UTE acquisitions and at both magnetic field strengths.

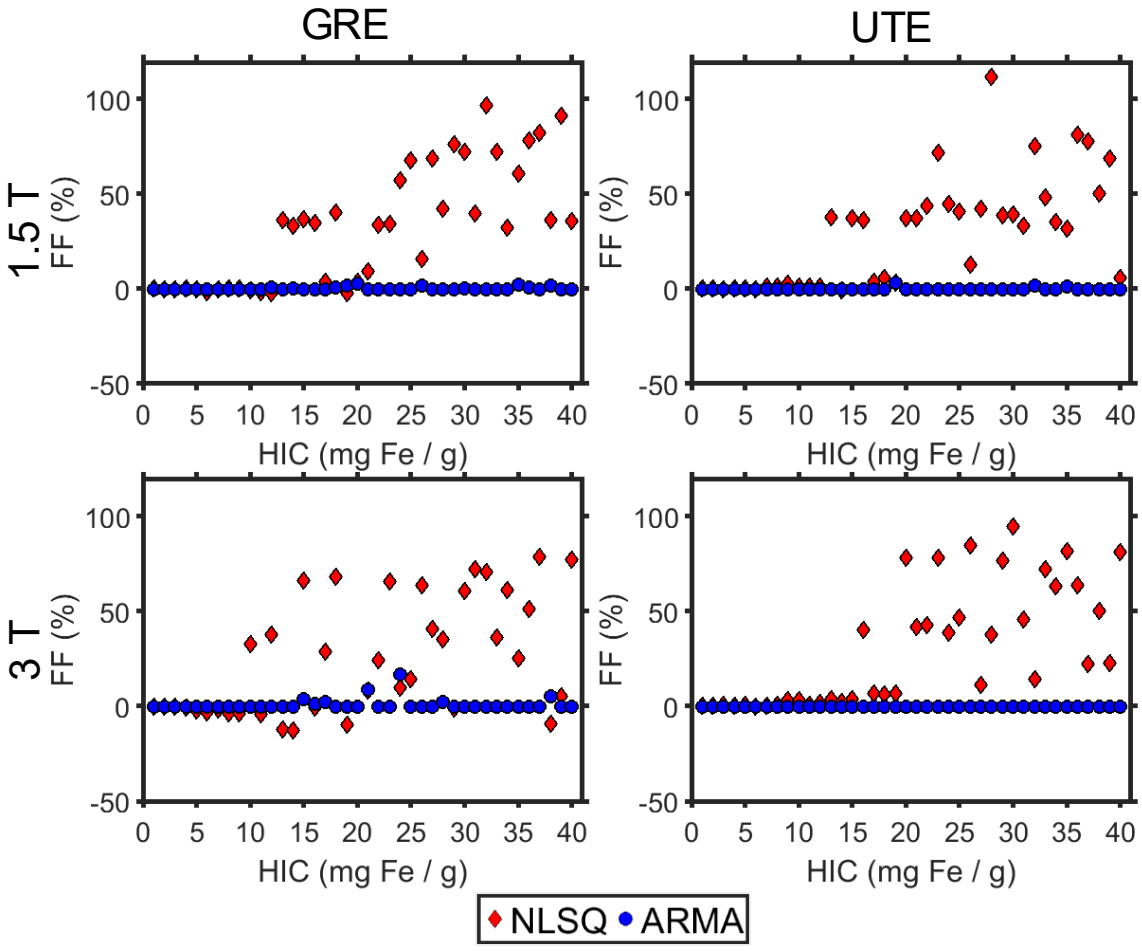

Figure 11. Mean FF (%) values estimated by NLSQ and ARMA models plotted against simulated HICs for GRE and UTE acquisitions (ΔTE = 0.5 ms, TEmax = 10 ms) at 1.5 T and 3 T. NLSQ model failed to estimate accurate FF content for GRE as well as UTE acquisition for HICs > 10 mg Fe/g. In contrast, ARMA model displayed true FF values close to zero throughout the HIC range for both GRE and UTE acquisitions.

Discussion

In this study, Monte Carlo-based virtual iron overload models were developed to mimic realistic human liver morphology and distributions, MRI signals were synthesized for different GRE and UTE acquisitions, and the R2* values estimated by monoexponential and multispectral fat-water models were analyzed. R2* results obtained using the monoexponential model showed an excellent agreement with the established in vivo R2*-HIC calibrations, illustrating the accuracy of our Monte Carlo-based simulation model. Our results demonstrate that both multispectral fat-water models, ARMA, and NLSQ showed high R2* accuracy and precision for UTE acquisition with shorter echo spacing ($\leq$ 0.5 ms) and longer TEmax ($\geq$ 6 ms) across the full clinical spectrum of HIC at both field strengths, with ARMA model producing similar results as the monoexponential model.

Most of the published iron simulation studies for R2* estimation generate synthetic MRI signals by just numerically varying the R2* values without considering the true dephasing effects of iron deposits on the simulated MRI signal.[29,33] As the size and distribution of iron deposits might have different dephasing effects on the MRI signal relaxation, simulating the true in vivo morphometry of iron deposition and modeling the iron-proton interactions will provide realistic MRI signal observed in liver iron overload. In this study, histologically realistic liver iron overload models incorporating the in vivo sizes and intrahepatocyte and interhepatocyte clustering of iron deposits were reproduced from Ghugre et al. study.[13,15] In our liver geometry, vascular sinusoids were not included and water protons were allowed to freely diffuse across hepatocyte boundaries as it was previously demonstrated that neither incorporating sinusoids, nor restricting proton diffusion alters R2* significantly.[13] Our results demonstrate that the synthesized MRI signals decayed faster with increasing HICs and at higher magnetic field strengths with R2* values

showing almost a two-fold increase at 3 T compared to at 1.5 T, which is in agreement with previous simulation and in vivo studies.[31,35]

The estimated R2* values obtained by fitting the monoexponential model to our simulated MRI signals at 1.5 T for GRE acquisition showed an excellent agreement with slopes close to published in vivo R2*-HIC calibrations. In contrast, multispectral fat-water signal models produced higher slopes, with the slope of the ARMA model still within the 95% confidence intervals of Wood et al. calibration while slope of NLSQ deviating outside of the confidence bounds for higher HICs. One reason for the discrepancies in slopes between multispectral fat-water models and published calibration studies that use monoexponential model may be the fact that fat-water models perform fitting on complex data whereas monoexponential model uses magnitude data for fitting. Due to this, both models differ in the noise consideration especially at high HICs for GRE acquisitions because of low signal-to-noise ratio (SNR), hence, potentially causing differences in R2* estimation.[33] Discrepancies in R2* estimation between signal models can strongly depend on the acquisition parameters chosen, particularly in terms of echo times.[29,36] Choosing the best set of echo times is crucial especially for higher HIC ranges for minimizing bias in model estimated R2* values.[35] Hence, in this study, we systematically investigated the performance of multispectral models for different GRE and UTE acquisition parameters to determine their accuracy and precision in R2* quantification across the full clinical range of iron overload.

Currently, multiecho GRE sequences are used for the diagnosis of iron overload in clinical practice.[24] However, our simulations demonstrated that all signal models exhibited poor R2* accuracy and precision for GRE acquisitions at higher HICs (> 20 mg Fe/g at 1.5 T, > 12.5 mg Fe/g at 3 T) similar to other studies due to rapid signal decay before the first TE of 1 ms.[10,12,29]

Further, the results did not significantly improve even using shorter ΔTE (≤ 0.5 ms) or shorter TEmax (≤ 4 ms) for GRE acquisitions at higher HICs and at 3 T for either of the models. In contrast, accuracy and precision in R2* measurements substantially improved for UTE acquisition (TE_1 = 0.1 ms) for both monoexponential and multispectral fat-water signal models similar to previous studies that used only monoexponential R2* model.[10,11,29] More importantly, our results show that the R2*-HIC slopes for monoexponential and multispectral models are much closer for UTE acquisitions at both field strengths for shorter ΔTE (≤ 0.5 ms), hence, demonstrating that UTE method can minimize noise bias and produce similar R2* results across the full range of HIC for both magnitude and complex fitting models.

Across the signal models, ARMA model showed similar R2* regression and precision trends as monoexponential model for GRE and UTE acquisitions when comparisons were made for the same ΔTEs, whereas NLSQ model showed comparatively more R2* deviation from linearity and lower precision than the other two models especially for ΔTE ≥ 1 ms. Decreasing TEmax from 10 ms to 6 ms while maintaining a ΔTE = 0.5 ms, did not have any significant impact on R2* accuracy or precision for all three signal models across the full HIC range for both GRE and UTE acquisitions. However, shortening TEmax to ≤ 4 ms caused bias in R2* estimation and produced higher CoVs for monoexponential and NLSQ models especially for lower HICs due to incomplete signal decay. The monoexponential model used in this study fits an additional constant parameter to account for noise floor but when shorter TEmax is used for lower HICs, the signal decay will not hit the noise floor and fitting the constant term can cause bias in R2* quantification.[29] Similarly, NLSQ model fits for additional parameters and not acquiring sufficient echoes (~ 6 ms) to capture enough signal decay can cause uncertainties in the fitting and hence in the R2* estimation.[35] In contrast, ARMA model seemed to be robust and produced similar R2*

values and CoVs even for shorter TEmax of 2 ms (3 echoes) and 4 ms (7 echoes), which is likely because ARMA behaves as a pure monoexponential model with no additional fitting terms in the absence of fat.[8] Nevertheless, all signal models produced high accuracy and precision in R2* values for UTE acquisition with $\Delta TE \leq 0.5$ ms and $TEmax \geq 6$ ms at both 1.5 T and 3 T across the full simulated HIC range.

NLSQ model displayed random inaccuracies in FF estimation for HICs > 10 mg Fe/g for both GRE and UTE acquisitions at 1.5 T and 3 T even though these iron models did not have fat. These results agree with a previous study that showed that NLSQ model overestimated FF in phantoms and patients with high iron overload (R2* > 500 s^{-1}) and no steatosis.[8] Further, based on the recent empirical relationship derived for the NLSQ multispectral model using Cramer Rao Lower Bound calculations,[9] FF estimates will be unreliable above R2* of ~473 s^{-1} at 1.5 T and 642 s^{-1} at 3 T for the GRE acquisition used in our study, which aligns with our simulation results. ARMA model, in contrast, consistently estimated FF values close to zero for both GRE and UTE acquisitions at 1.5 T and 3 T because it only accounts for fat in the model when fat peak is detected, otherwise it behaves as a pure monoexponential model.[8]

There are some notable limitations to this study. First, our study relied solely on simulations and the R2* estimation by the signal models was not validated using phantom or patient data. However, performing MRI scans with various acquisition methods and parameters is quite challenging and expensive. In contrast, the simulation approach in this study can be applied for testing any acquisition technique and signal model while accounting for the true in vivo morphology of iron overload and helps to investigate the limits of the multispectral signal models for R2* quantification before applying them to real data. Another limitation of this study is that the multispectral signal models are not investigated in co-existing conditions of steatosis, which

will need creating realistic iron overload and steatosis virtual liver models mimicking true in vivo morphology,[37] and thereby, requires future studies for thorough investigation. Further, the R2* accuracy and precision of signal models are not investigated under different SNR conditions, and this warrants future investigation as SNR can influence the performance of the signal models.[29,33]

In conclusion, our study reproduced virtual iron overload models mimicking in vivo morphology, synthesized MRI signals and investigated the R2* accuracy and precision of multispectral fat-water models in comparison to monoexponential model for GRE and UTE acquisitions for various ΔTEs and TEmax values. In comparison to published in vivo R2*-HIC calibrations, monoexponential model produced similar slopes while ARMA and NLSQ models produced higher slopes for the GRE acquisition. All signal models produced similar R2*-HIC slopes for UTE acquisition for shorter echo spacing (≤ 0.5 ms) and longer TEmax (≥ 6 ms) across the full clinical spectrum of HICs at both 1.5 T and 3 T, with monoexponential and ARMA models producing similar R2* results. However, NLSQ model produced false FF values for high iron overload conditions (HICs ≥ 10 mg Fe/g) for both GRE and UTE acquisitions. Future work involves simulating liver models in coexisting conditions of iron overload and steatosis and investigating the performance of multispectral models for various acquisition methods for accurate quantification of R2* and fat.

Chapter 5 - Validation of Simulation Results Using Phantoms and Patient Data

Introduction

Liver biopsy is the conventional method used for diagnosing iron overload, but it has some drawbacks such as being invasive, painful, and having poor sampling variability.[34] Fortunately, magnetic resonance imaging (MRI) has emerged as a reliable alternative for assessing iron stores in the liver of patients with iron overload. Noninvasive MRI-based biomarkers, such as effective transverse relaxation or R2*, have shown promise as a quantitative biomarker with high diagnostic accuracy for quantifying hepatic iron concentrations (HICs).[5,34]

Currently, in vivo R2* vs HIC calibrations rely on a Cartesian-based multiecho gradient echo (GRE) sequence for imaging and a monoexponential model for fitting.[5] However, when fat and iron coexist in the liver, the MRI signal no longer follows a purely monoexponential decay and instead exhibits oscillations due to the difference in resonance frequency between the fat and water peaks.[8] To address this, researchers are exploring multispectral fat-water-R2* models, such as the NLSQ and ARMA models, which can recognize both water and fat peaks simultaneously, to achieve a more robust simultaneous quantification of both R2* and FF.[8,26] While some studies have shown that both NLSQ and ARMA models start giving unreliable estimates of R2* at high iron overload when using GRE sequences, center out radial sampling-based UTE sequences have been found to expand the dynamic range of R2* quantification due to their short initial echo (~ 0.1 ms).[29] However, there is a need for a systematic and thorough investigation of multispectral NLSQ and ARMA signal models in both UTE and GRE sequences.

To test the effects of the choice of signal models on R2*, researchers have been using MRI testing alternatives like phantoms and simulations that mimic the presence of hepatic iron or fat or both.[34] While Monte Carlo simulations can make conducting research more feasible, the

results still need to be validated using phantoms and patient data from MRI exams before being used clinically. Therefore, the proposed study aims to validate the reliability of data collected using simulations by using MRI images of phantoms as well as patients with varying degrees of iron overload, scanned using GRE and UTE sequences. By validating the results obtained from simulations to those obtained from phantom and patient data, the proposed study will provide a comprehensive understanding of the performance of multispectral fat-water-R2* models and help pave the way for their eventual clinical use.

Methods

Phantom

Ten cylindrical iron phantoms were created using 2 % agar-water mixtures and Magnefy iron particles. The phantoms had various iron concentrations (0, 0.0075, 0.025, 0.04, 0.06, 0.1, 0.2, 0.4, 0.8, 1.4) (%) and no fat to cover clinically relevant R2* values. The phantoms were arranged in a 2 x 5 rectangular array and scanned at 1.5 T in the Diagnostic Imaging Department at St Jude Children's Research Hospital in Memphis, TN (Avanto, Siemens Healthineers, Malvern, PA). Unfortunately, unavailability of source code for UTE at 3 T prevented scanning the phantoms at this magnetic strength. The sequences at 1.5 T were chosen so that the echo times matched those that yielded the best results in the simulation study. For the 1.5 T scan, the acquisition parameters were as follows: GRE: TE_1 = 1.2 ms, ΔTE = 1.44 ms, TEmax = 9.84 ms, 7 echoes, monopolar readout gradient, matrix = 128 x 104, flip angle = 25º, slice thickness = 5 mm; UTE: TE_1 = 0.1 ms, ΔTE = 0.5 ms, TEmax = 9.6 ms, 17 echoes, 5 interleaves, matrix = 192 x 192, flip angle = 20º, slice thickness = 10 mm. The MRI images were obtained in DICOM format and analyzed using MATLAB (R2020b, The MathWorks, Inc., Natick, MA, USA). Ten circular ROIs were drawn manually on the phantom images, and R2* maps were calculated

using monoexponential with noise subtraction, ARMA, and NLSQ fitting models.[6] Time constraints did not allow the development of monoexponential with constant offset model used in simulation studies to be optimized accurately for phantom and patient data.[5] The performance of the fitting models was evaluated using regression analysis between the mean R2* ± SD values and Fe concentrations, and by studying the corresponding R2* maps. Finally, FF maps were also analyzed to determine the accuracy of the multispectral models in estimating the absence of fat in the iron-only phantoms.

Figure 12. Vials of 10 phantoms of different iron concentrations arranged in a 2 * 5 rectangular array for imaging.

Patients

The in vivo data used in this study were obtained from an institutional review board (IRB) approved retrospective study conducted at the Diagnostic Imaging Department at St Jude Children's Research Hospital, from patients diagnosed with transfusional iron overload and scanned between April 2013 and February 2017 using Avanto, Siemens Healthineers scanners. Due to limited availability of data, only three datasets each at 1.5T and 3 T were analyzed for

this study. Datasets were chosen such that the parameters used for imaging were comparable to those used in our simulation study. The imaging parameters for the patient data were as follows:

For GRE:

At 1.5 T: Monopolar readout gradient, matrix = 128 x 104, TE_1 = 1.2 ms, ΔTE = 1.44 ms, TEmax = 9.84 ms, 7 echoes, slice thickness = 5 mm, flip angle = 45° for Patient 1, TE_1 = 1.15 ms, ΔTE = 1.72 ms, TEmax = 9.75 ms, 6 echoes, slice thickness = 5 mm, flip angle = 45° for Patient 2, TE_1 = 1.07 ms, ΔTE = 1.38 ms, TEmax = 9.35 ms, 7 echoes, slice thickness = 5 mm, flip angle = 45° for Patient 3.

At 3 T: Monopolar readout gradient, matrix = 128 x 104, TE_1 = 1.04 ms, ΔTE = 1.29 ms, TEmax = 8.78 ms, slice thickness = 5 mm, flip angle = 35° for Patient 4, TE_1 = 1.1 ms, ΔTE = 1.48 ms, TEmax = 9.98 ms, slice thickness = 5 mm, flip angle = 20°, for both Patient 5 and 6.

For UTE at both 1.5 T and 3 T:

TE_1 = 0.1 ms, ΔTE = 0.5 ms, TEmax = 9.6 ms, 17 echoes, 5 interleaves, slice thickness = 10 mm, flip angle = 20°, matrix = 192 x 192 for all six patients.

Quantitative R2* maps were calculated using the magnitude-based monoexponential with noise subtraction, and complex domain-based fat-water-R2* NLSQ and ARMA models. Mean R2* values of the whole liver without blood vessels were extracted by applying a ROI mask, obtained by thresholding pixel values via histogram analysis (Figure 13). Finally, mean liver R2* values were converted to HIC estimates for each model using a previously published R2*-HIC biopsy calibration,[6] and were compared for accuracy with the corresponding biopsy values obtained from the study conducted in St. Jude Children's Research Hospital. FF maps were also analyzed

to determine the accuracy of the multispectral models in estimating the absence of fat in the patients diagnosed with iron overload only in our study.

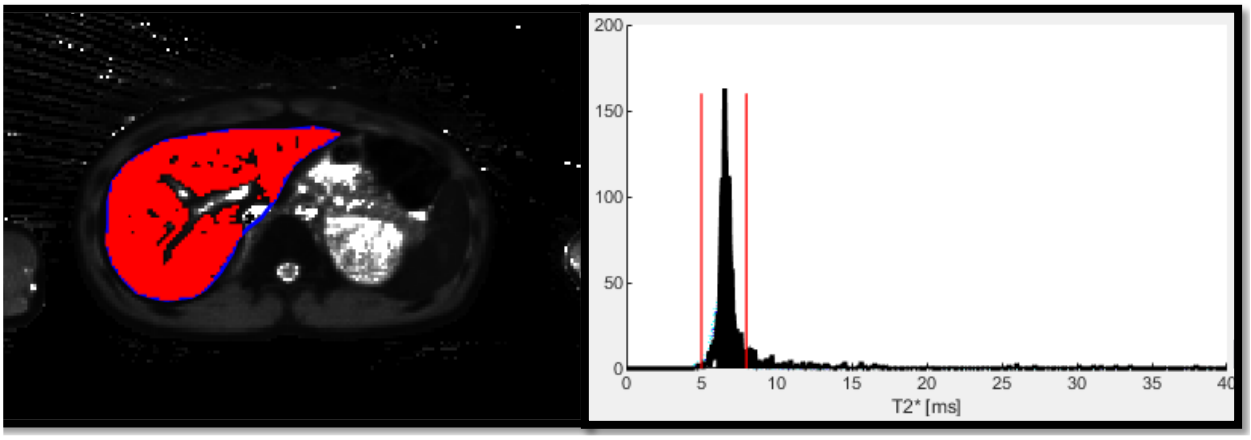

Figure 13. Manually selected ROI covering the whole liver area, excluding blood vessels based on histogram analysis for a representative patient scanned using UTE acquisition at 1.5 T.

Results

Phantoms

Figure 14 illustrates the results of regression analysis between Fe concentrations of Magnefy phantoms and mean R2* ± SD values estimated using monoexponential, ARMA, and NLSQ signal models for GRE and UTE acquisitions at 1.5 T magnetic strength. For GRE sequence, all three models failed to maintain linearity at higher iron concentrations of R2* greater than ~1500 s^{-1} and exhibited high SDs. When compared to the GRE sequence, the 1.5 T UTE acquisition produced superior results for all three models, throughout the full range of Fe concentrations and reached beyond the 1500s^{-1} range that models using GRE acquisition could not quantify. The monoexponential and ARMA models showed similar mean R2* values, SD, and linearity trends for all Fe concentrations. Although the NLSQ model exhibited a comparatively greater SDs at the higher Fe concentrations and underestimated R2* for the highest Fe concentration (1.4 %), its performance significantly improved for the UTE acquisition compared to GRE. Analysis of the R2* maps revealed inhomogeneous regions at higher iron concentrations that did not increase linearly in all three models for GRE acquisitions (Figure 15). Maps at UTE acquisitions, on the other hand, revealed all three signal models demonstrated linearly increase in R2* values with iron concentrations, with some inhomogeneous region at the highest iron concentration for NLSQ model. Finally, FF maps revealed that NLSQ model was estimating false FF values for extreme iron concentration in some of the phantoms obtained with both GRE and UTE acquisitions (Figure 16). On the other hand, ARMA model accurately estimated FF values close to 0 % for all phantoms for both acquisitions.

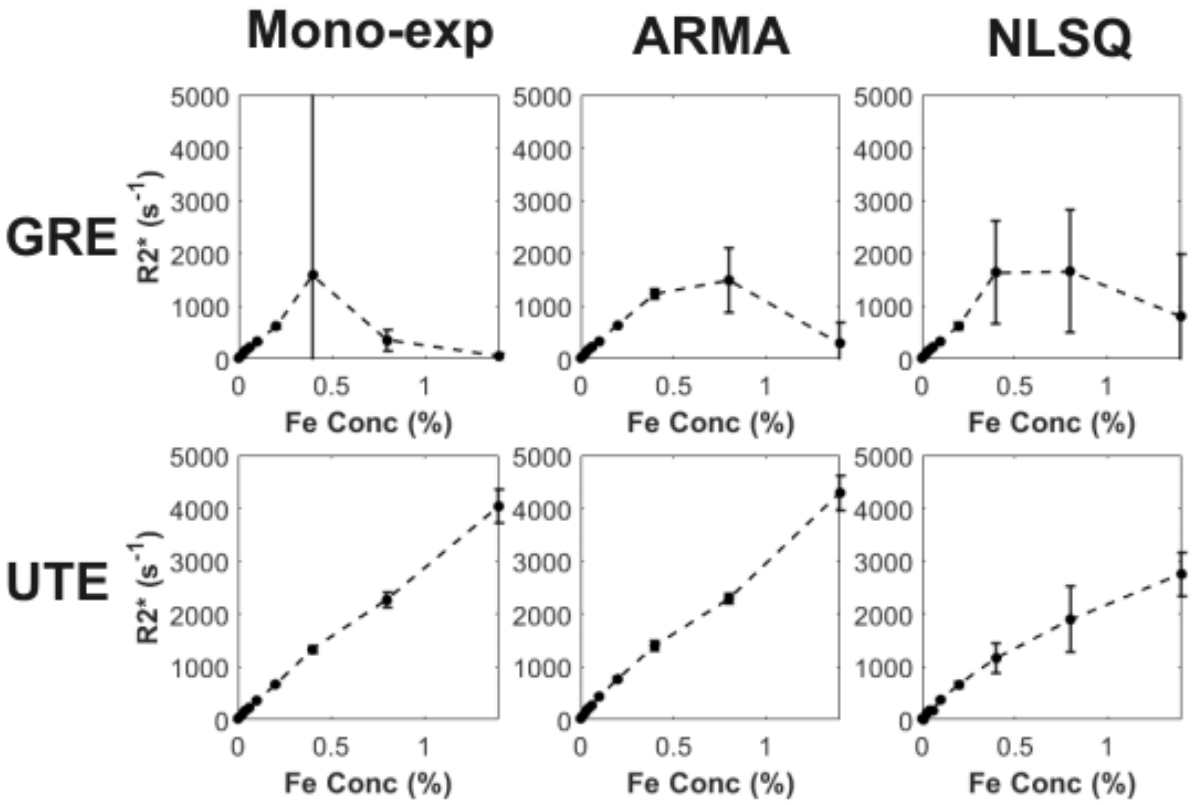

Figure 14. Regression analysis between Fe concentrations of Magnefy phantoms and mean R2* ± SD values estimated using monoexponential, ARMA, and NLSQ signal models for GRE and UTE acquisitions at 1.5 T. Performance of monoexponential, ARMA, and NLSQ models significantly improved for UTE acquisition compared to GRE acquisition, with monoexponential and ARMA models performing similarly for UTE acquisition.

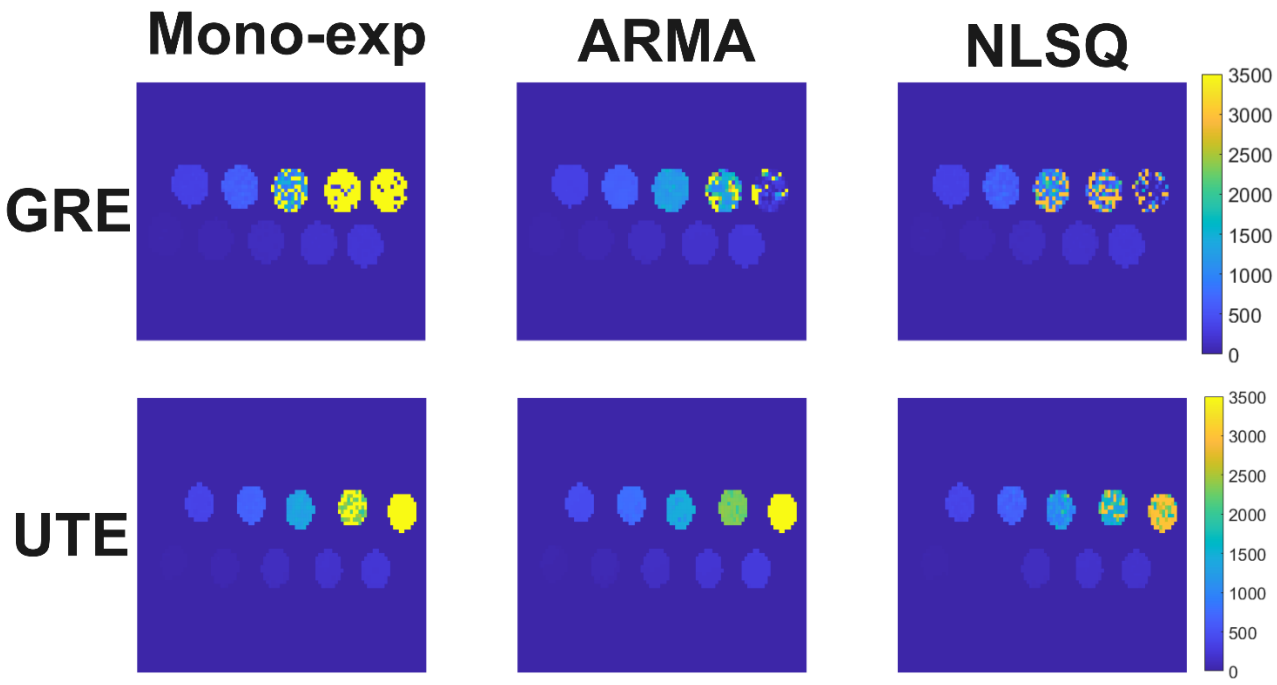

Figure 15. Monoexponential, ARMA, and NLSQ R2* maps of the 10 vials of phantoms with different iron concentrations scanned at 1.5 T for GRE and UTE acquisitions. Performance of all three signal models significantly improved at UTE acquisition compared to GRE acquisition, with monoexponential and ARMA signal models exhibiting similar R2* maps.

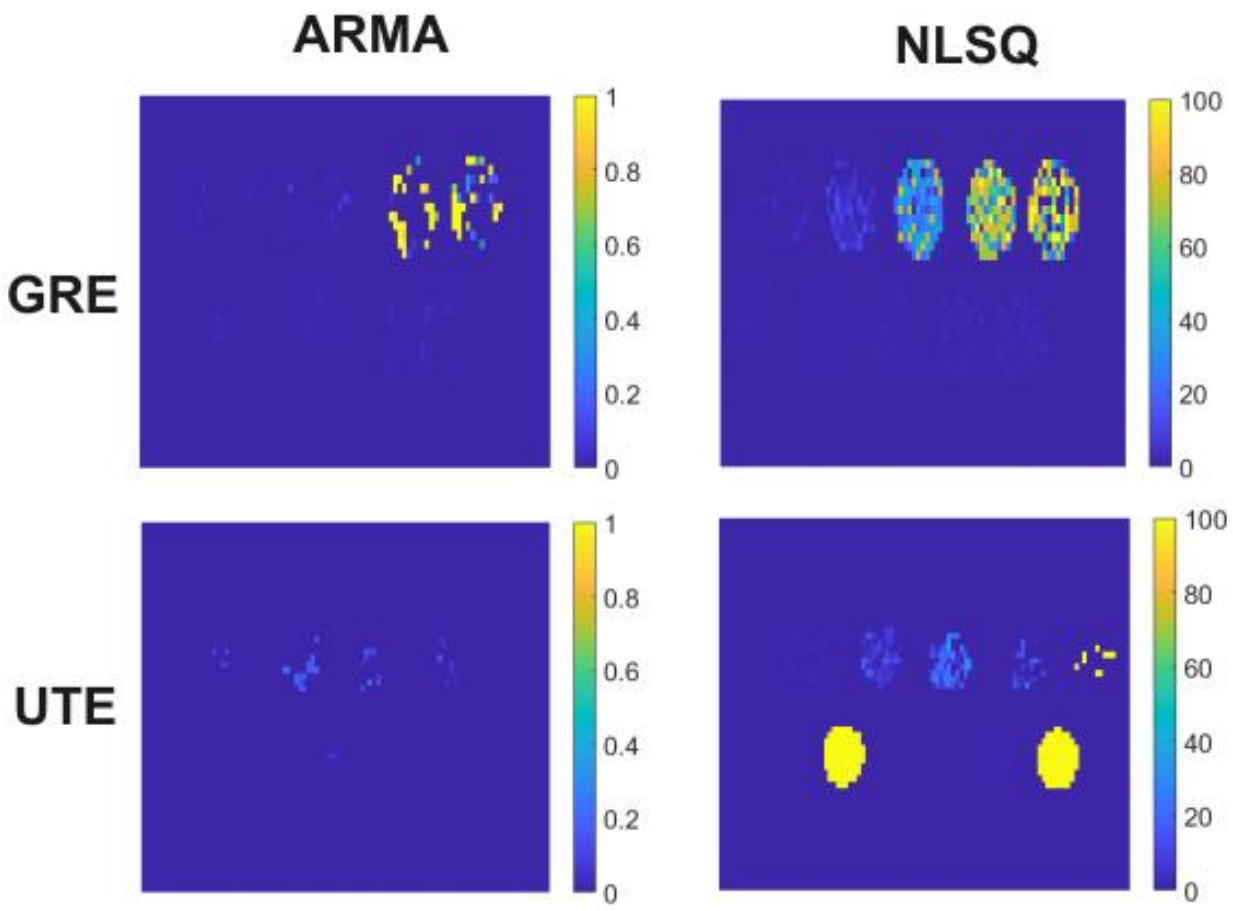

Figure 16. FF maps of ARMA and NLSQ signal models for GRE and UTE acquisitions at 1.5 T. ARMA model displayed FF % close to zero for all iron phantoms whereas NLSQ model displayed inconsistent values of higher FF % for some phantoms.

Patients

Figure 17 and 18 show the R2* maps calculated using monoexponential, ARMA, and NLSQ models for GRE and UTE acquisitions at 1.5 T and 3 T, respectively. The monoexponential model estimated HIC exhibited excellent agreement with true biopsy HIC for all patients at both 1.5 T and 3 T for both GRE and UTE acquisitions, with some degrees of deviations from biopsy HICs and inhomogeneous regions in the R2* maps, especially for Patient 4 diagnosed with a higher HIC (Figure 18). ARMA signal model estimated HIC values in agreement with biopsy

values for GRE acquisitions at both magnetic strengths in all patients, with some degrees of deviations. However, the degree of deviation from biopsy HIC increased relatively for ARMA model for UTE acquisitions at both 1.5 T and 3 T, especially with an increase in HIC. Additionally, NLSQ model estimated HIC values in agreement with biopsy values for both GRE and UTE acquisitions at mild iron overload for both magnetic strengths. However, NLSQ model exhibited higher deviations from biopsy HIC for patients with higher HIC (Patient 2 and 4), for both GRE and UTE acquisitions. Overall, monoexponential and ARMA signal model were more in agreement for GRE acquisition than UTE, for both fields. FF maps of all patients revealed that ARMA model exhibited FF values close to 0 % for all degrees of iron overload used in our study, whereas NLSQ model displayed varying degrees of false FF %. Figure 19 shows a representative example of FF maps of Patient 2 scanned at 1.5 T.

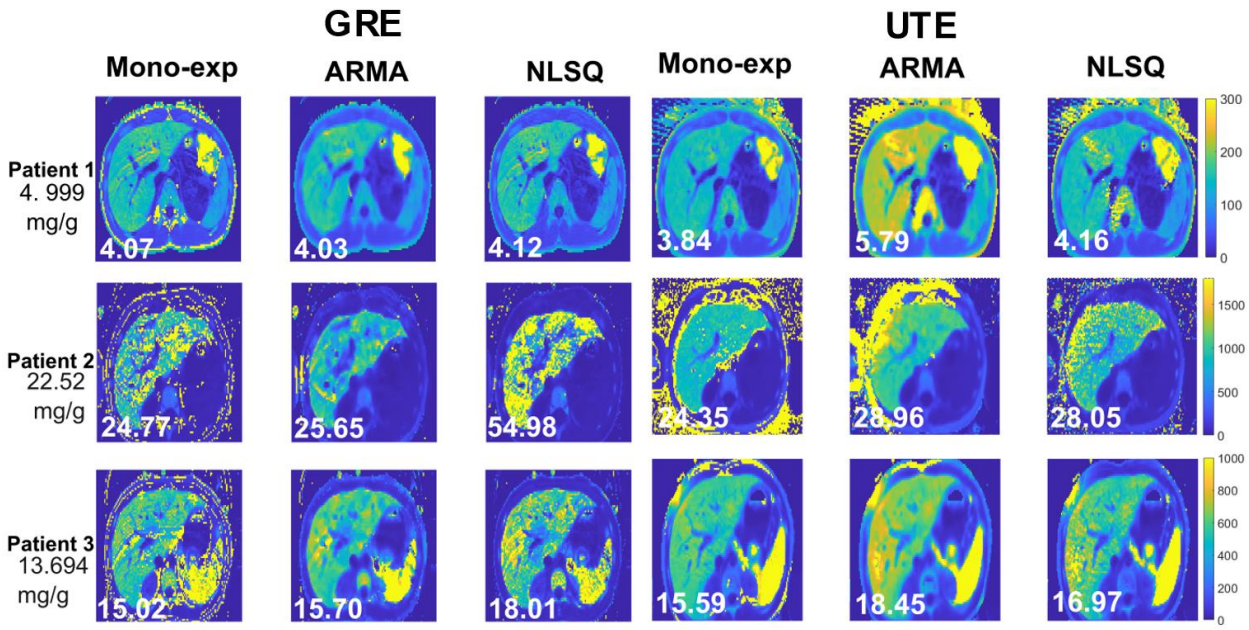

Figure 17. Monoexponential, ARMA, and NLSQ R2* maps of the three patients scanned at 1.5 T for both GRE and UTE acquisitions, with corresponding biopsy-derived HIC values and model-estimated HIC values (provided on the side and R2* maps, respectively). For GRE acquisition, the monoexponential and ARMA models estimated HIC values that were in better agreement with the biopsy-derived values for all patients. Deviation from biopsy HIC increased for UTE acquisition using ARMA model and for higher HICs using NLSQ model.

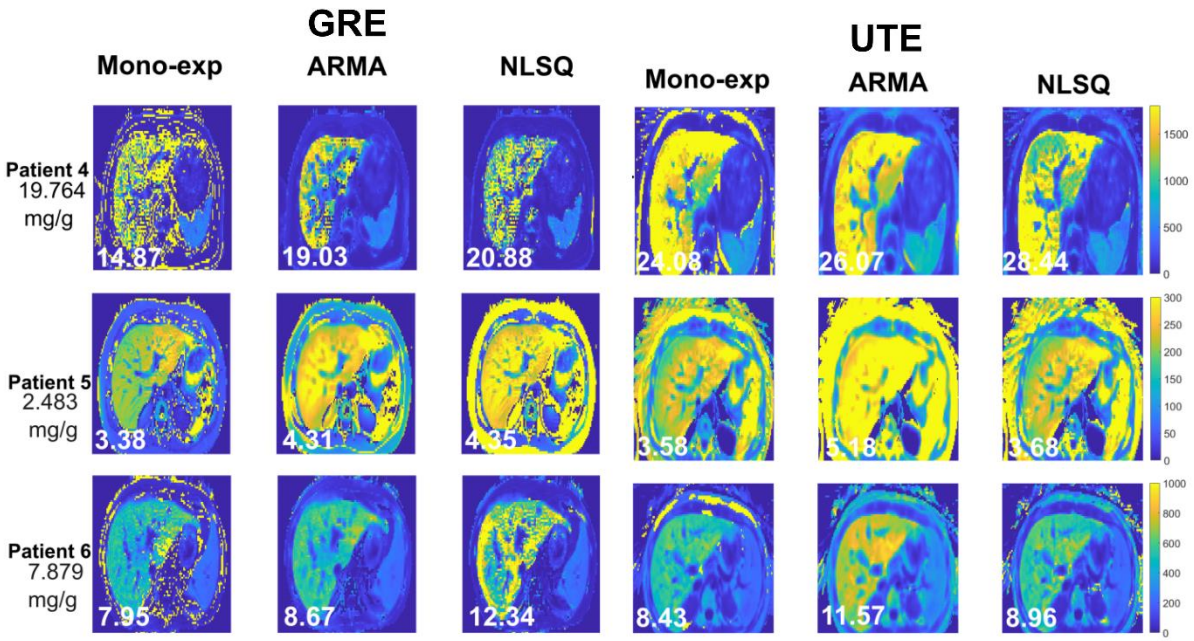

Figure 18. Monoexponential, ARMA, and NLSQ R2* maps of the three patients scanned at 3 T for both GRE and UTE acquisitions, with corresponding biopsy-derived HIC values and model-estimated HIC values (provided on the side and R2* maps, respectively). The monoexponential and ARMA models for GRE acquisition showed better agreement with biopsy-derived HIC for all patients. Deviation from biopsy HIC increased for UTE acquisition when using the ARMA model and for higher HICs when using the NLSQ model.

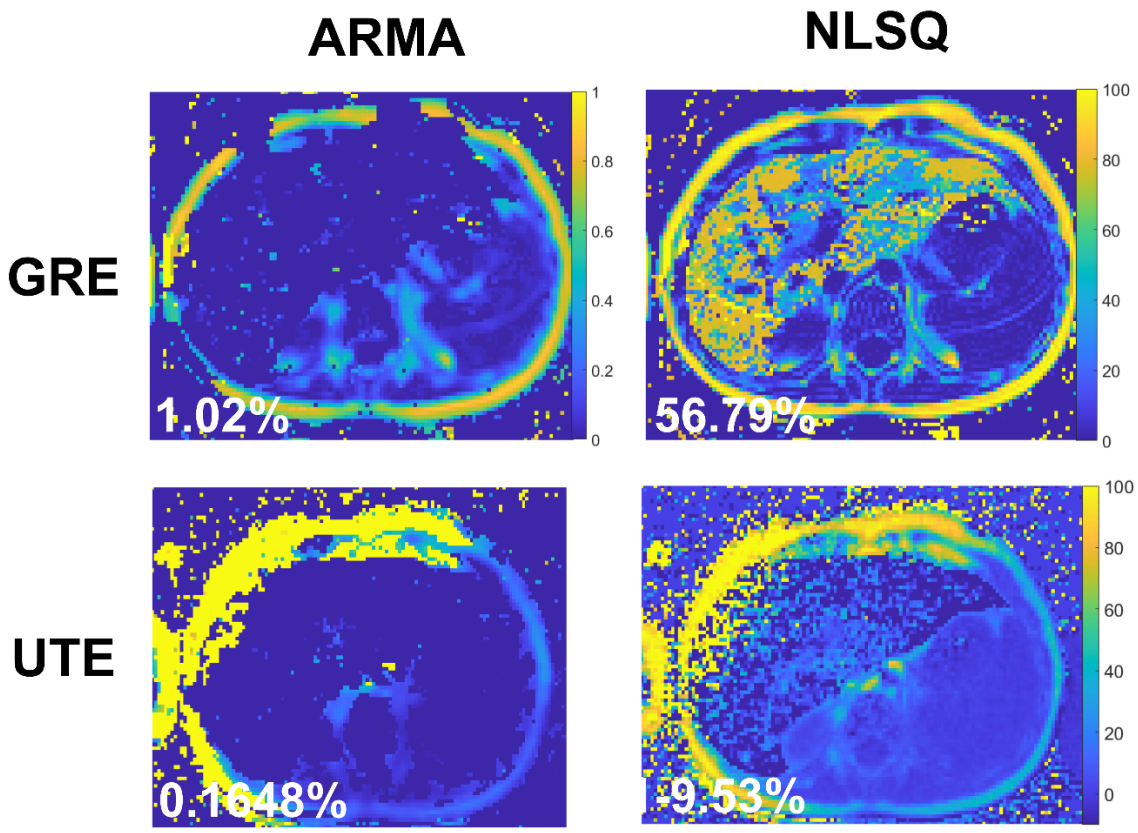

Figure 19. FF map of representative Patient 2 (Biopsy HIC = 22.52 mg/g, Steatosis description = No evidence of steatosis) scanned at 1.5 T for both GRE and UTE acquisitions. ARMA model estimated FF close to 0 for both GRE and UTE acquisitions, whereas NLSQ model estimated inconsistent values of false FF %.

Discussion

This study aimed to validate the results of simulation studies using phantoms and patient data to evaluate multispectral fat-water-R2* models for GRE and UTE acquisitions. Results from the phantom analysis revealed that all three signal models showed a significant improvement in performance for UTE acquisition with short ΔTE (≤ 0.5 ms) and longer TEmax (≥ 6 ms) compared to GRE acquisition with a longer TE_1, validating the conclusions from the Monte Carlo simulation study. Additionally, similar to the findings from simulations, the monoexponential and ARMA signal models performed similarly for UTE acquisitions compared to the NLSQ signal model. Results from the patient study showed that the monoexponential and ARMA signal models provided accurate HIC estimations for GRE acquisitions. However, results for UTE acquisition demonstrated decreased reliability for all patients for the ARMA model; ARMA method potentially failed due to inconsistent echo spacing used for the UTE sequence.

In the phantom study, the R2* values were underestimated for iron concentrations $\geq 0.4\%$ for all three signal models at GRE acquisition, with larger SDs. It is worth noting that these high iron concentrations used in the study appeared very dark on the MRI images, which could have affected the accuracy of the GRE sequence with $TE_1 \sim 1$ ms in quantifying their fast-decaying MRI signals (refer to the Appendix).[8] Upon examination of their calculated R2* maps, it was observed that these iron phantoms (0.4, 0.8, 1.4 %) appeared to be more inhomogeneous in all three models, which seemed to have adversely affected the R2* values. In contrast to GRE, the use of center-out radial sampling method in UTE imaging, resulted in a significant improvement in the performance for all three models. Our findings are consistent with other research studies that have also demonstrated UTE acquisition to extend the dynamic range of HIC estimation at 1.5 T.[10,29]

The patient study did not validate the results from our simulation study. The monoexponential model accurately estimated HIC for all patients in all acquisitions, with some degrees of discrepancies with the biopsy values and inhomogeneous regions in the R2* maps. To identify the root cause of these discrepancies and inhomogeneities, further examination of the accuracy of R2* maps and signal estimates used in the model is required, especially for Patient 4 at 3 T.[5,13] The performance of the ARMA signal model was similar to that of the monoexponential model for the GRE acquisition, as it accurately estimated HIC values for all six patients, with some deviation from biopsy HIC. However, for the UTE acquisition, the ARMA model estimated HIC with greater discrepancy from biopsy HIC compared to the GRE acquisition and did not show the similarity trends seen in simulations and phantoms with the monoexponential model. In contrast, the NLSQ model performed better than the ARMA model overall for the UTE acquisition for our patient study. However, the NLSQ model also did not validate the results from the simulation, as the estimated HICs did not closely match biopsy HICs for UTE acquisition for patients with high HIC. Upon examination of these results, several possible reasons were identified for the discrepancies observed. Firstly, the ARMA signal model assumes the echoes to be equidistant, but the UTE sequence used in our study had five interleaves with unequal echospacings, which may have contributed to the discrepancies observed.[8] Secondly, the UTE sequence used in this study was a UTE fat suppressed sequence that is used to suppress signals coming from the subcutaneous regions of fat in the abdomen.[11,29] However, the use of this fat suppressed sequence to decrease the streaking artifacts in the MRI images could have caused inaccuracies in the R2* estimation for the NLSQ and ARMA signal models.[29] Lastly, the low SNR in patients with high iron overload and rapidly decaying signals

may have contributed to the lack of sufficient signals for accurate HIC estimation with the multispectral models.

In line with the findings of our simulation studies and a prior study by Tipirneni-Sajja et al,[8] the NLSQ model estimated false fat fractions in our phantom and patient study as well. As previously discussed, the NLSQ model relies on prior knowledge about signal peaks and frequencies, which can sometimes result in inaccurate fat fraction measurements.[8,28] This is because the model may mistakenly identify fat peaks where none exist, leading to false fat fraction estimates.[8] On the other hand, since ARMA signal model identifies individual signal peaks, it behaves as a pure monoexponential model in the absence of fat.[8]

Our study had several limitations that should be addressed in future research. First and foremost, we made several attempts to optimize the monoexponential with constant offset model used in our simulation study, but we were unable to achieve satisfactory results on the MRI images within the given time frame. As a result, we resorted to using the monoexponential with noise subtraction model proposed by Hankins et al.[6] The difference between the two fitting models lies in their algorithm to account for fat; the monoexponential with constant offset assumes the MRI signal is the sum of a monoexponential decay and a constant offset due to background noise or other factors.[5,37,38] On the other hand, monoexponential with noise subtraction model assumes that the noise in the data is random and uses regression analysis to estimate R2* value from the noise-subtracted signal.[6,29,37] Therefore, it is important to consider that the monoexponential models used in the simulation study, and the patient and phantom study are not the same, when validating the results. Moreover, while liver biopsy is considered the gold standard for measuring iron overload, it is still subject to sampling variability since it does not provide a complete view of the liver.[34] Therefore, the HIC values estimated from biopsy

results may not be completely accurate. It is important to note that the models used in our study estimate HIC values based on whole liver analysis, which may result in some degree of discrepancy when compared to the biopsy-derived values. One important factor to note is the presence of ghosting and aliasing artifacts due to breathing motion in some of the images at 1.5 T GRE, which may have affected R2* estimation by the models. To determine if artifacts influenced our results, a larger sample size of patients should be studied to assess the reproducibility of our findings. Another limitation is the difference in FOV between the GRE and UTE images, which may have introduced bias into our results. Due to technical difficulties and time constraints, we were unable to obtain phantom and patient data in the manner that we had initially planned. Future studies should use sequences with comparable FOV and echo spacing parameters to minimize this source of error.

Future research should incorporate 3 T scanned phantoms to gain a more complete understanding of the potential impact of acquisitions on signal models. It is also obvious that the extreme SD for the 0.4 % phantom at 1.5 T GRE is unrealistic, so the vial of phantom, images, and iron particle used, could be reinvestigated. Lastly, ARMA and monoexponential signal models have the advantage of faster signal processing when compared to the NLSQ model. This highlights one of their strengths over the NLSQ model. Future research can focus on exploring the reasons behind the slower performance of the NLSQ model and identifying potential solutions to address this issue. Additionally, the performance of the ARMA signal model can be optimized in the future by using UTE sequences with equidistant echoes or modifying the model to account for interleaves.

Overall, this study is a work in progress and further data collection and analysis is necessary to draw definitive conclusions about the performance of signal models for assessing iron overload in phantoms and patient data.

Chapter 6 – Conclusion and Future Work

As noninvasive approaches to assessing and treating diseases become more attractive and preferred, extensive research is being done on MRI techniques for evaluating iron overload. The current R2*-HIC calibrations, obtained using a combination of multiecho gradient echo (GRE) imaging and monoexponential fitting, work well for iron-only conditions,[5,6] but a more robust combination of acquisition parameters and signal models is needed to account for the presence of fat in the liver.[8,26] The multispectral fat water models are being tested in different acquisition settings of GRE and UTE imaging to extend the feasibility of quantitative MRI studies.

Simulation and phantom results showed that iron overload models accurately mimic human liver morphology and realistic MRI signals can be synthesized for different acquisitions and field strengths that exhibit greater relaxivity with increasing iron conentrations. Both multispectral ARMA and NLSQ signal models had improved R2* accuracy and precision for UTE acquisition with shorter echo spacing and longer TEmax across the full clinical spectrum of HIC at both field strengths, with ARMA and monoexponential signal models performing more similarly than NLSQ. However, patient data did not yield definitive results for UTE acquisition, and future work should validate the simulation results using comparable UTE sequences for patients across the full range of iron overload.

Appendices

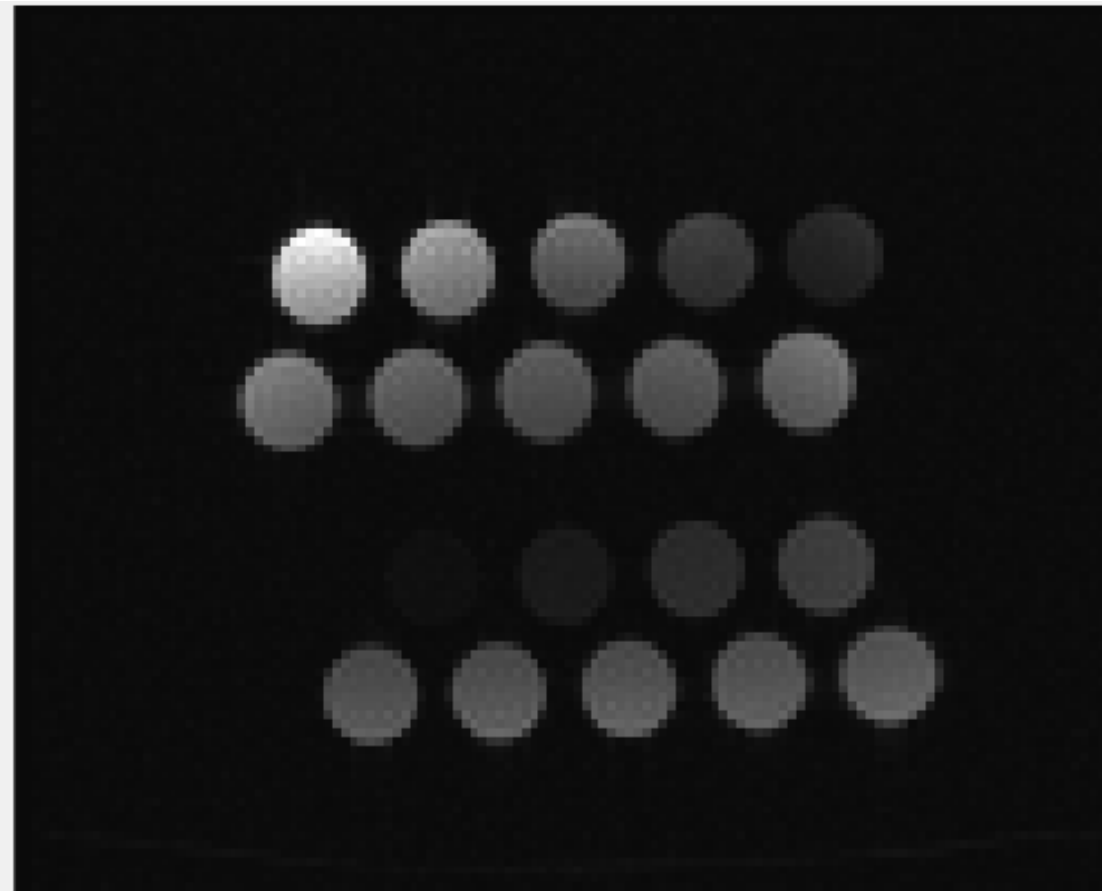

Figure 20. The MRI image shown here displays 10 vials of Magnefy phantoms scanned at 1.5 T (with the top two rows featuring phantoms that were not used in our study and can be disregarded). Notably, the 0.4%, 0.8%, and 1.4% phantoms appear very dark in the images, indicating very high iron concentrations and helps explain why GRE sequence with a longer TE_1 may have been inadequate for accurately quantifying the R2* values.

References

1. Porter JL, Rawla P. Hemochromatosis. In: *StatPearls.* Treasure Island (FL)2022.
2. Brittenham GM, Badman DG, National Institute of D, Digestive, Kidney Diseases W. Noninvasive measurement of iron: report of an NIDDK workshop. *Blood.* 2003;101(1):15-19.
3. McDowell LA, Kudaravalli P, Sticco KL. Iron Overload. In: *StatPearls.* Treasure Island (FL)2022.
4. Henninger B. Demystifying liver iron concentration measurements with MRI. *Eur Radiol.* 2018;28(6):2535-2536.
5. 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.
6. 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.
7. 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.* 2015;187(6):472-479.
8. Tipirneni-Sajja A, Krafft AJ, Loeffler RB, et al. Autoregressive moving average modeling for hepatic iron quantification in the presence of fat. *J Magn Reson Imaging.* 2019;50(5):1620-1632.
9. Colgan TJ, Zhao R, Roberts NT, Hernando D, Reeder SB. Limits of Fat Quantification in the Presence of Iron Overload. *J Magn Reson Imaging.* 2021;54(4):1166-1174.
10. Doyle EK, Toy K, Valdez B, Chia JM, Coates T, Wood JC. Ultra-short echo time images quantify high liver iron. *Magn Reson Med.* 2018;79(3):1579-1585.
11. Krafft AJ, Loeffler RB, Song R, et al. Quantitative ultrashort echo time imaging for assessment of massive iron overload at 1.5 and 3 Tesla. *Magn Reson Med.* 2017;78(5):1839-1851.

12. Kee Y, Sandino CM, Syed AB, et al. Free-breathing R 2 * mapping of hepatic iron overload in children using 3D multi-echo UTE cones MRI. *Magn Reson Med.* 2021;85(5):2608-2621.
13. Ghugre NR, Wood JC. Relaxivity-iron calibration in hepatic iron overload: probing underlying biophysical mechanisms using a Monte Carlo model. *Magn Reson Med.* 2011;65(3):837-847.
14. Ghugre NR, Doyle EK, Storey P, Wood JC. Relaxivity-iron calibration in hepatic iron overload: Predictions of a Monte Carlo model. *Magn Reson Med.* 2015;74(3):879-883.
15. Ghugre NR, Gonzalez-Gomez I, Shimada H, Coates TD, Wood JC. Quantitative analysis and modelling of hepatic iron stores using stereology and spatial statistics. *J Microsc.* 2010;238(3):265-274.
16. Labranche R, Gilbert G, Cerny M, et al. Liver Iron Quantification with MR Imaging: A Primer for Radiologists. *Radiographics.* 2018;38(2):392-412.
17. Sirlin CB, Reeder SB. Magnetic resonance imaging quantification of liver iron. *Magn Reson Imaging Clin N Am.* 2010;18(3):359-381, ix.
18. Ganz T. Hepcidin and its role in regulating systemic iron metabolism. *Hematology Am Soc Hematol Educ Program.* 2006:29-35, 507.
19. Fleming RE, Britton RS, Waheed A, Sly WS, Bacon BR. Pathophysiology of hereditary hemochromatosis. *Semin Liver Dis.* 2005;25(4):411-419.
20. Wallace DF. The Regulation of Iron Absorption and Homeostasis. *Clin Biochem Rev.* 2016;37(2):51-62.
21. 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.
22. 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.
23. Deugnier Y, Turlin B. Pathology of hepatic iron overload. *World J Gastroenterol.* 2007;13(35):4755-4760.

24. Flora SJ, Pachauri V. Chelation in metal intoxication. *Int J Environ Res Public Health.* 2010;7(7):2745-2788.
25. Nielsen P, Engelhardt R, Dullmann J, Fischer R. Non-invasive liver iron quantification by SQUID-biosusceptometry and serum ferritin iron as new diagnostic parameters in hereditary hemochromatosis. *Blood Cells Mol Dis.* 2002;29(3):451-458.
26. Hernando D, Liang ZP, Kellman P. Chemical shift-based water/fat separation: a comparison of signal models. *Magn Reson Med.* 2010;64(3):811-822.
27. Taylor BA, Loeffler RB, Song R, McCarville MB, Hankins JS, Hillenbrand CM. Simultaneous field and R2 mapping to quantify liver iron content using autoregressive moving average modeling. *J Magn Reson Imaging.* 2012;35(5):1125-1132.
28. Hernando D, Kellman P, Haldar JP, Liang ZP. Robust water/fat separation in the presence of large field inhomogeneities using a graph cut algorithm. *Magn Reson Med.* 2010;63(1):79-90.
29. Tipirneni-Sajja A, Loeffler RB, Krafft AJ, et al. Ultrashort echo time imaging for quantification of hepatic iron overload: Comparison of acquisition and fitting methods via simulations, phantoms, and in vivo data. *J Magn Reson Imaging.* 2019;49(5):1475-1488.
30. Xia Y. *Essential concepts in MRI : physics, instrumentation, spectroscopy and imaging.* First edition. ed. Hoboken, NJ: John Wiley & Sons, Inc.; 2022.
31. Wang C, Reeder SB, Hernando D. Relaxivity-iron calibration in hepatic iron overload: Reproducibility and extension of a Monte Carlo model. *NMR Biomed.* 2021;34(12):e4604.
32. Yamada I, Aung W, Himeno Y, Nakagawa T, Shibuya H. Diffusion coefficients in abdominal organs and hepatic lesions: evaluation with intravoxel incoherent motion echo-planar MR imaging. *Radiology.* 1999;210(3):617-623.
33. Horng DE, Hernando D, Reeder SB. Quantification of liver fat in the presence of iron overload. *J Magn Reson Imaging.* 2017;45(2):428-439.

34. 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.
35. 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.
36. 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.
37. Shrestha U, van der Merwe M, Kumar N, et al. Morphological characterization of hepatic steatosis and Monte Carlo modeling of MRI signal for accurate quantification of fat fraction and relaxivity. *NMR Biomed.* 2021;34(6):e4489.