MAGNETIC RESONANCE IMAGING OF SKELETAL INFLAMMATION: A QUANTITATIVE APPROACH

Timothy JP Bray

University College London

A thesis submitted to University College London for the degree of Doctor of Philosophy

2018
I, Timothy Bray, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.
Abstract

‘Spondyloarthritis’ is an umbrella term referring to a group of inflammatory diseases characterized by spinal inflammation and new bone formation, which cause pain, disability and reduced quality of life. Magnetic resonance imaging (MRI) is commonly used to identify, quantify and monitor inflammation in patients with spondyloarthritis, and can therefore help to guide treatment aimed at reducing inflammation. However, conventional methods for interpretation of MRI scans are qualitative, lack reproducibility and provide only indirect information about tissue pathology. Therefore, there is a need for a more objective method for identifying and quantifying inflammation using MRI. Quantitative MRI (qMRI) enables objective physical measurements of tissue characteristics to be made directly from the image, and is a candidate tool. In this thesis, two main qMRI techniques - diffusion-weighted imaging and chemical shift-encoded MRI, which derive apparent diffusion coefficient (ADC) and proton density fat fraction (PDFF) measurements respectively - have been considered as potential methods for quantifying inflammation. ADC measurements are known to increase in areas of bone marrow oedema; here, it was shown that ADC measures reflect response to treatment in patients undergoing tumour necrosis factor inhibitor (TNFi) therapy. CSE-MRI was optimized as a new method for imaging inflammation, and new fat-water-bone phantoms were designed enabling technical validation of PDFF measurements. PDFF was compared in areas of bone marrow oedema, fat metaplasia and normal marrow, and was shown to decrease in oedematous sites and increase in areas of fat metaplasia. A new partially-automated tool for measuring ADC and PDFF measurements in the subchondral bone using histographic analysis was developed, and histographic parameters were compared in a prospective study of 53 patients. Histographic analysis was shown to improve the performance of both ADC and PDFF measurements for identifying inflammation and fat metaplasia. Additionally, potential methods for quantifying bone mineral density (BMD) – and thus quantifying bone formation and bone loss – were evaluated. $R_2^*$ and quantitative susceptibility measurements were shown to reflect bone mineral density (BMD) in fat-water-bone phantoms, and also showed differences in areas of fat metaplasia compared to normal bone marrow. The results confirm the feasibility of using qMRI to quantify inflammation in spondyloarthritis. The fat-water-bone phantom and quantification tools described here could be used in future studies aiming to quantify and characterize bone inflammation.
Acknowledgements

The content of this thesis is the result of collaborative work with a number of individuals who I would like to acknowledge for their support and assistance, without which the project would not have been completed.

First, I would like to thank Margaret Hall-Craggs, who gave me the opportunity to undertake this project and who has been a constant source of support, encouragement and ideas over the last three years. She has provided an incredibly supportive environment which has allowed me to explore and to develop as a researcher.

I am extremely grateful to John Ioannou for sharing his expertise on the rheumatological aspects of the project and for his help with the design of the studies. I would also like thank Alan Bainbridge, who has been a huge help with the physics aspects of the research and has taught me a great deal about magnetic resonance imaging. Shonit Punwani has provided valuable advice and support regarding the imaging, and this work could not have been completed without Andre Lopes’ input on statistical design.

I would also like to thank Debajit Sen, Lucy Wedderburn and Corinne Fisher for supporting the project throughout, Ania Radziszewska and Hannah Peckham for their help with serum biomarkers, and Ania Karsa and Karin Shmueli for their help with susceptibility mapping. This work would not have been possible without all of those who have volunteered for the studies contained here, of whom there are too many to name. I would also like to thank my friends at CMI for making my time so enjoyable.

Last but definitely not least, I am grateful to my wife Siyamini, my new daughter Sahana and my family (Mum, Dad, George and Chris) for loving and supporting me throughout this whole process.
Academic Output

The following is a list of works published by the author during the course of the doctorate.

**Peer-reviewed Journal Articles**


Singh S, **Bray TJP**, Hall-Craggs MA. Quantifying bone structure, micro-architecture and pathophysiology with MRI. Clin Rad 2018;73(3):221-230


†Indicates equal contributions

**Peer-reviewed Conference Abstracts (selected)**


**Bray TJP**, Vendhan K, Atkinson D, Fisher C, Sen D, Ioannou Y, Hall-Craggs MA. Quantification of Sacroiliac Joint Inflammation using Diffusion-weighted imaging in
Young People: Biological Validation in Enthesitis-related Arthritis. International Congress on Spondyloarthritides 2016, Ghent, Belgium.


Walters U, Fisher C, Bray TJP, Hall-Craggs MA, Sen D. Sacroiliac joint imaging in adolescent back pain: Discrepancy between clinical features and MRI findings. Rheumatology 56 (supplement 6).


Contents

ABSTRACT ................................................................................................................................. 5
ACKNOWLEDGEMENTS ................................................................................................................ 7
ACADEMIC OUTPUT .................................................................................................................... 9
CONTENTS .................................................................................................................................. 14
LIST OF FIGURES ....................................................................................................................... 20
LIST OF TABLES .......................................................................................................................... 28
LIST OF ABBREVIATIONS ........................................................................................................... 30
1 OVERVIEW ............................................................................................................................... 31
1.1 THE CLINICAL NEED .............................................................................................................. 31
1.2 ORGANISATION OF THE THESIS ......................................................................................... 32
1.3 AIMS AND OBJECTIVES ....................................................................................................... 34
2 INTRODUCTION ......................................................................................................................... 35
2.1 OVERVIEW ............................................................................................................................ 35
2.2 SPONDYLOARTHRITIS AND THE NEED FOR IMAGING ...................................................... 35
  2.2.1 The Spondyloarthritides ..................................................................................................... 35
    2.2.1.1 Overview ..................................................................................................................... 35
    2.2.1.2 Epidemiology .............................................................................................................. 37
    2.2.1.3 Clinical features and disease course ........................................................................... 37
    2.2.1.4 Genetics and pathophysiology .................................................................................. 39
    2.2.1.5 Histology .................................................................................................................. 43
    2.2.1.6 Spondyloarthritis in children and adolescents ............................................................ 48
    2.2.1.7 Relationship between age at onset and prognosis ...................................................... 51
    2.2.1.8 Treatment of spondyloarthritis .................................................................................. 52
    2.2.1.9 Serum biomarkers in spondyloarthritis ...................................................................... 54
  2.2.2 Why Image? – and the Case for MRI ................................................................................ 55
    2.2.2.1 Diagnosis .................................................................................................................... 56
    2.2.2.2 Measuring disease activity ......................................................................................... 59
    2.2.2.3 Prediction of therapeutic response ............................................................................. 60
    2.2.2.4 Prediction of sustained remission ............................................................................... 60
    2.2.2.5 Prediction of phenotype and severity ......................................................................... 61
    2.2.2.6 Monitoring structural changes ................................................................................... 61
    2.2.2.7 Limitations of MRI .................................................................................................... 62
2.3 INTRODUCTION TO MAGNETIC RESONANCE IMAGING

2.3.1 Introduction

2.3.2 Nuclear Magnetic Resonance

2.3.3 Image Formation

2.3.4 Creating Echoes

2.3.5 Diffusion-weighted Imaging

2.3.6 Chemical Shift-encoded MRI

2.4 MRI METHODS FOR QUANTIFYING INFLAMMATION

2.4.1 Conventional MRI

2.4.2 STIR Images Show Bone Marrow Oedema

2.4.3 T1-weighted Images Provide ‘Structural’ Information

2.4.4 Quantitative Methods for Imaging Inflammation
3 THE APPARENT DIFFUSION COEFFICIENT AS A BIOMARKER OF INFLAMMATION

3.1 OVERVIEW ........................................................................................................113
3.2 BACKGROUND ................................................................................................113
3.3 THE LINEAR ROI METHOD ...........................................................................114
  3.3.1 Rationale .....................................................................................................114
  3.3.2 Description of Method ..............................................................................115
3.4 ADC AS A MARKER OF RESPONSE TO THERAPY ......................................117
  3.4.1 Introduction ................................................................................................117
  3.4.2 Methods .....................................................................................................117
    3.4.2.1 Subjects ...............................................................................................117
    3.4.2.2 Image Acquisition .............................................................................118
    3.4.2.3 Image Analysis ..................................................................................118
    3.4.2.4 Response Classification .....................................................................119
    3.4.2.5 Statistical Analysis ............................................................................120
  3.4.3 Results .........................................................................................................120
    3.4.3.1 Demographics ....................................................................................120
    3.4.3.2 Disease Response: Radiological Classification ....................................121
    3.4.3.3 Disease Response: Clinical Classification ..........................................121
    3.4.3.4 Receiver operating characteristic analysis .........................................124
    3.4.3.5 ΔADC, ΔnADC as Continuous Response Measures .........................125
    3.4.3.6 Inter-observer Variability ...................................................................126
  3.4.4 Discussion ....................................................................................................126
3.5 CONCLUSION ....................................................................................................129
3.6 ACKNOWLEDGEMENTS ..................................................................................130

4 RATIONALE FOR USING CSE-MRI AND METHOD OPTIMISATION .............133

4.1 OVERVIEW .......................................................................................................133
4.2 RATIONALE FOR USING CSE-MRI TO IMAGE BONE INFLAMMATION .......133
4.3 OPTIMISATION OF CSE-MRI FOR IMAGING BONE MARROW ...................134
  4.3.1 Overview ....................................................................................................134
  4.3.2 Evaluation of Candidate Methods ...............................................................135
    4.3.2.1 IDEAL ..................................................................................................135
    4.3.2.2 Magnitude-based Reconstruction .......................................................137
    4.3.2.3 ‘Berglund + magnitude’ ......................................................................143
    4.3.2.4 Commercial CSE-MRI Packages .......................................................146
5 PDFF AND R2* AS INFLAMMATORY BIOMARKERS: PROOF-OF-PRINCIPLE ... 151
  5.1 INTRODUCTION ........................................................................................................ 151
  5.2 METHODS .................................................................................................................. 151
    5.2.1 Fat-Water and Fat-Water-Bone Phantoms ......................................................... 151
    5.2.2 Patients and Volunteers ...................................................................................... 152
    5.2.3 Imaging Sequence and Reconstruction .............................................................. 155
    5.2.4 Statistical Analysis ............................................................................................ 155
  5.3 RESULTS ...................................................................................................................... 156
    5.3.1 Fat-Water and Fat-Water-Bone Phantoms ......................................................... 156
    5.3.2 Patients and Volunteers ...................................................................................... 159
    5.3.3 Repeatability ....................................................................................................... 161
  5.4 DISCUSSION ............................................................................................................... 161
  5.5 CONCLUSION ............................................................................................................. 166
  5.6 ACKNOWLEDGEMENTS ............................................................................................ 166

6 THE BEACH TOOL: COMPARISON OF ADC AND PDFF AS INFLAMMATORY
BIOMARKERS IN A PROSPECTIVE STUDY ................................................................. 167
  6.1 INTRODUCTION ........................................................................................................ 167
  6.2 METHODS .................................................................................................................. 168
    6.2.1 Patient Selection ................................................................................................ 168
    6.2.2 Image Acquisition ............................................................................................. 168
    6.2.3 Definition of Regions-of-Interest ....................................................................... 169
    6.2.4 Histographic Analysis ....................................................................................... 170
    6.2.5 Scoring of Conventional MRI .......................................................................... 173
    6.2.6 Clinical Scores .................................................................................................... 173
    6.2.7 Statistical analysis ............................................................................................. 174
  6.3 RESULTS ...................................................................................................................... 174
    6.3.1 Inflammatory Parameters .................................................................................. 174
    6.3.2 Structural Parameters ....................................................................................... 177
    6.3.3 Relationship between qMRI parameters and Visual Scores .............................. 178
    6.3.4 Relationship between MRI and Symptoms ...................................................... 179
    6.3.5 Interobserver Agreement .................................................................................. 180
  6.4 DISCUSSION ............................................................................................................... 180
  6.5 CONCLUSION ............................................................................................................. 183
Contents

6.6 SUPPLEMENTARY INFORMATION ...........................................................................184
  6.6.1 Generation of polygonal ROIs ........................................................................184
  6.6.2 Derivation of PDFF and ADC Thresholds .........................................................185

7 QUANTITATIVE SUSCEPTIBILITY MAPPING FOR MEASUREMENT OF BONE

7.1 INTRODUCTION .....................................................................................................187

7.2 METHODS ...............................................................................................................188
  7.2.1 Phantoms ........................................................................................................188
  7.2.2 Patients and Volunteers ..................................................................................189
  7.2.3 Data Acquisition ............................................................................................189
  7.2.4 Derivation of Quantitative Susceptibility Maps from Phase Images ...........189
  7.2.5 Statistical Analysis ..........................................................................................193

7.3 RESULTS ...............................................................................................................194
  7.3.1 Phantoms ........................................................................................................194
  7.3.2 Patients and volunteers ...................................................................................194

7.4 DISCUSSION ..........................................................................................................201

7.5 CONCLUSION .........................................................................................................204

8 DISCUSSION AND PERSPECTIVES .........................................................................205

8.1 OVERVIEW ............................................................................................................205

8.2 DIRECTIONS FOR FURTHER RESEARCH .............................................................207
  8.2.1 Improved Imaging of Oedema .........................................................................207
  8.2.2 The Varying Composition of Normal Bone .....................................................209
  8.2.3 Measuring Bone Formation and Destruction with MRI ..............................210
  8.2.4 Whole Body Imaging for Measuring the Inflammatory Burden ....................212
  8.2.5 ‘Functional’ Imaging of Inflammation? ............................................................214

8.3 TOWARDS CLINICAL IMPLEMENTATION ............................................................215

8.4 CONCLUSION .........................................................................................................217

9 APPENDIX A ............................................................................................................219

9.1 PHYSICAL BASIS OF MR SIGNAL ........................................................................219
  9.1.1 Conversion between laboratory and rotating Frames ......................................219
  9.1.2 MR signal during frequency encoding .............................................................219
  9.1.3 MR signal during phase encoding ....................................................................220

9.2 CSE-MRI ALGORITHMS .......................................................................................220
  9.2.1 IDEAL ..............................................................................................................220
### Contents

9.2.2 Magnitude-based fitting ................................................................. 222
9.2.3 ROI propagation for BEACH Tool ..................................................... 224
  9.2.3.1 Createpoly .................................................................................. 224
  9.2.3.2 Use of createpoly to define two polygonal ROIs ............................. 226

10 APPENDIX B ......................................................................................... 227

10.1 ASSOCIATION OF THE APPARENT DIFFUSION COEFFICIENT WITH MATURITY IN ADOLESCENT SACROIILAC JOINTS .......................................................... 227
  10.1.1 Introduction .................................................................................... 227
  10.1.2 Methods ......................................................................................... 227
  10.1.3 Results .......................................................................................... 231
  10.1.4 Discussion ...................................................................................... 233
  10.1.5 Acknowledgements ........................................................................ 235

10.2 DIAGNOSTIC UTILITY OF WHOLE BODY DIXON MRI IN MULTIPLE MYELOMA – A MULTI-READER STUDY ........................................................................................................... 235
  10.2.1 Introduction .................................................................................... 235
  10.2.2 Methods ......................................................................................... 236
  10.2.3 Results .......................................................................................... 242
  10.2.4 Discussion ...................................................................................... 248
  10.2.5 Conclusion ..................................................................................... 250
  10.2.6 Acknowledgements ........................................................................ 251

REFERENCES ............................................................................................. 253
List of Figures

Figure 2-1 - Classification systems for spondyloarthritis. a. The traditional classification system emphasizes differences between subtypes. b. The modern ASAS classification system categorises spondyloarthritides into broader groups depending on axial or peripheral involvement; the axial group is further subdivided depending on the presence of ‘radiographic’ spondyloarthritis.

Figure 2-2 - Histological image of inflamed bone marrow after staining for IL6 (a pro-inflammatory cytokine). There are a large number of inflammatory cells (arrow), with a smaller number of adipocytes (labelled fat), in the bone marrow. Reproduced with permission from (66) under License Number 4179260946789.

Figure 2-3 - Histological images showing bone marrow in control (A) and AS patient (B). The iliac bone marrow (solid line) is normal in the control but has a ‘myxoid’ appearance in AS due to the presence of an inflammatory exudate. Reproduced with permission from (66) under License Number 4179260946789.

Figure 2-4 - Invasion of the joint space by fibrovascular granulation tissue (solid arrowhead arrow, top). The cartilage (labelled C) appears to be being eroded by adjacent mononuclear inflammatory cells (open/v-shaped arrowhead, bottom). Reproduced with permission from (65) under License Number 4179280903311.

Figure 2-5 - Trabecular changes in iliac bone in spondyloarthritis. A shows an 18 year old female control, and B shows a patient with AS. On the iliac side of the bone (arrowed), the end plate is thinned and the subchondral bone is eroded, whilst the adjacent para-articular bone is sclerotic. L denotes the ligamentous joint, S denotes the sacrum and D denotes the synovial (diarthrodial) joint. Reproduced with permission from (65) under License Number 4179280903311.

Figure 2-6 - TNF inhibitors. Fv - variable region, Fc - fragment crystallisable region (interacts with immune system).

Figure 2-7 - SPARC SCoring system. Each sacroiliac joint is divided into four quadrants, and the observer allocates a point for each quadrant where inflammation is present. Extra points are given if the inflammation (solid arrow) is at least as intense as presacral blood vessels (dashed arrow), or if the signal changes extend more than 1cm from the joint surface.

Figure 2-8 - Precession. The axis of the magnetic moment rotates around the magnetic field.

Figure 2-9 - Behaviour of magnetisation during an RF pulse. At the start of the pulse (t=0), the magnetisation is aligned in the direction of the positive z-axis. During the pulse (at time t), the magnetization M(t) is gradually tipped in the yz plane (rotating frame), with flip angle $\theta$. At the end of the inversion pulse (t=T), the magnetisation is aligned in the direction of the negative z-axis.

Figure 2-10 - A graphical representation of relaxation (simulated data in MATLAB) in the laboratory frame.
Figure 2-11 - Slice selection. The RF pulse and slice selection gradient are applied simultaneously, enabling selective excitation of a specific tissue slice. 73

Figure 2-12 - Frequency encoding. 74

Figure 2-13 - Frequency encoding. The spatial frequencies being 'interrogated' increase throughout the frequency encode gradient, such that the highest spatial frequency data is acquired last. 75

Figure 2-14 - k-space. Each phase encoding step (two individual phase encoding steps are labelled in red and blue) corresponds to a specific row in k-space. Each element corresponds to a specific pair of spatial frequencies – i.e. specific values for kFE and kPE, as defined in Appendix A (Section 9.1) A 2D Fourier transform is used to reconstruct the image. 77

Figure 2-15 - Gradient echo sequence. 78

Figure 2-16 - Spin echo sequence. 80

Figure 2-17 - Relationship between tumbling rate and relaxation time. 81

Figure 2-18 - Spin-spin relaxation and T2 contrast. Contrast between tissues with varying T2 can be optimized by an appropriate choice of TE, shown by the dashed grey line. The data shown were simulated using MATLAB. 82

Figure 2-19 - Spin-lattice relaxation and T1 contrast. Contrast between tissues with varying T1 can be optimized by an appropriate choice of TR, as shown by the dashed grey line. The data shown were simulated using MATLAB. 83

Figure 2-20 - Fat saturation using CHESS. 85

Figure 2-21 - The STIR sequence for fat suppression. 87

Figure 2-22 - Schematic illustration showing displacement distributions of diffusing molecules. The variance of the Gaussian distribution increases over time. 89

Figure 2-23 - The Stejskal Tanner sequence. 90

Figure 2-24 - Fat-water ambiguity. In mixed voxels (a, containing both fat and water) interference at opposed phase echo times leads to loss of signal. In pure water (b) and pure fat (c) voxels, the magnitude of the signal is unchanged at opposed phase echo times, leading to ambiguity between pure water and pure fat, as in Dixon's original method. However, phase data can be used to distinguish between these situations (i.e. resolve the fat-water ambiguity). 93

Figure 2-25 - Fat and water signal vectors without field inhomogeneity (a), and with field inhomogeneity included (b). When field inhomogeneity is added the phase angle (labelled \( \theta \)) between fat and water remains constant, and both vectors are subject to an additional phase shift (labelled \( \phi \) depending on \( fB \). 95

Figure 2-26 - Single and multipeak fat models. In the multipeak model there are a number of fat peaks (shown schematically), each with a different chemical shift value, as shown in the Table below. This leads to additional complexity in the behaviour of the signal, as shown in Figure 2-27, below. 98
Figure 2-27 – NMR spectrum of subcutaneous fat (left) and tibial bone marrow (right), measured by Ren et al. (201). The bottom trace shows the acquired spectrum, and the upper trace shows the fitted spectrum. 99

Figure 2-28 – Simulation of signal evolution in the presence of spectral complexity of fat and T2* decay. Simulated signal/time plots are shown for (a) Single peak fat model with no T2* decay (b) Single peak fat model with T2* decay (c) Complex spectral fat model with no T2* decay and (d) Complex spectral fat model with T2* decay. The fat spectrum described by Ren et al. (201) was used for these simulations. 99

Figure 2-29 - Extensive, bilateral, active inflammation (bone marrow oedema) of the sacroiliac joint in a patient with spondyloarthritis. There is high signal on the T2-weighted STIR image (a), low signal on the T1-weighted image (b), and high signal on a fat-saturated T1-weighted image after gadolinium injection (c). 107

Figure 2-30 - Chronic inflammatory ('structural') lesions in spondyloarthritis. (a,b): Bilateral erosions of the sacroiliac joints (solid red arrow) are shown, surrounded by areas of fat metaplasia (blue arrow). (c,d): There is bilateral ankylosis of the sacroiliac joints (open red arrow) and extensive fat metaplasia (blue arrow). 110

Figure 3-1 - Linear ROI technique. Placement of linear regions of interest (ROIs) (red) is shown on the apparent diffusion coefficient (ADC) map of a patient with sacroiliitis (a) and a control with mechanical back pain (b). Only one slice is shown for each patient (four slices were scored in total). (c) The reference ROI is placed on interforaminal sacral bone (blue line) and used to generate a reference profile. (d) ADC profiles from the two subjects (in red) are shown compared with the reference profile from interforaminal sacral bone (in blue). The ADC values are higher for the patient with sacroiliitis (left) than for the normal subject (right). 116

Figure 3-2 - Preliminary results showing nADC measurements against visual inflammation scores. 116

Figure 3-3 - Response plots for nADC and ADC by radiological response. Pre- and post-treatment nADC and ADC values are shown for all 22 patients, classified according to radiological response. Patients whose nADC/ADC reduced between after treatment are shown in green, whilst patients whose nADC/ADC increased are shown in red. ADC values have units mm²/s x 10⁻⁶. 122

Figure 3-4 - Response plots for nADC and ADC by clinical response. Pre- and post-treatment nADC and ADC values are shown for all 22 patients, classified according to clinical response. Patients whose nADC/ADC reduced between after treatment are shown in green, whilst patients whose nADC/ADC increased are shown in red. ADC values have units mm²/s x 10⁻⁶. 123

Figure 3-5 - ROC analysis. 125

Figure 3-6 - ΔADC and ΔnADC as continuous response measures. 125
Figure 3-7 - Effect of fat metaplasia on the apparent diffusion coefficient. Areas of fat metaplasia (solid arrow) and normal marrow (dashed arrow) are shown on T1-weighted (a) and STIR (b) images, and ADC maps (c). ADC values are lower in areas of fat metaplasia than in normal marrow. ................................................................. 129

Figure 3-8 - Fat ghosting artifacts in DWI. Two representative slices (A and B) are shown from a single patient. For each slice, b=0 images (top row) and ADC maps (bottom row) are included. (a,b): Fat ghosts cover the right ilium, and would therefore alter quantitative measurements. (c,d): Switching the phase encoding direction to RL alters the location of the artifacts and would therefore minimize their impact. However, as shown in (g, h), this switch can cause artifacts in other slices. ................................................................. 131

Figure 4-1 - Example fat fraction maps from the two subjects' datasets, generated using IDEAL. ........................................................................................................................................... 136

Figure 4-2 - Acetone-water phantom. The percentages shown refer to the acetone fraction by volume; adjusted acetone fractions were used for the analysis. ........................................................................................................ 138

Figure 4-3 - Results of magnitude fit. Note that all fat-dominant voxels (e.g. subcutaneous fat) appear to be water-dominant due to unresolved fat-water ambiguity – i.e. they are 'swapped'. .................................................................................. 140

Figure 4-4 - Signal oscillations in five of the eleven tubes in the acetone phantom. ....................... 140

Figure 4-5 - ω plotted against acetone fraction over the range of acetone fractions in the phantom. In all the tubes containing a mixture of water and acetone (i.e. acetone fraction 0.1 to 0.9), there was a gradual reduction in ω with increasing acetone fraction. ................. 141

Figure 4-6 - Effect of flip angle on measured acetone fraction. ............................................................ 142

Figure 4-7 - ‘Berglund + magnitude’ method. .................................................................................... 144

Figure 4-8 - Examples of images from Berglund and 'Berglund + magnitude' methods. ................. 145

Figure 4-9 - Fat fraction and R2* maps generated using Philips mDixon Quant. ......................... 148

Figure 4-10 - Agreement between measured PDFF and reference fat fraction for mDixon Quant and 'Berglund + Magnitude' methods. .................................................................................. 149

Figure 5-1 - Fat-water (a,b) and fat-water-bone (c,d) phantoms. a: 0%, 30% and 70% FF vials are shown lying flat to demonstrate solidity. b: Microscopic images of 50% FF tube shows the arrangement of fat and water in the emulsion (the fat globules are circular and and outlined by dark lines). c: Photograph of the fat-water-bone phantom, showing fat content varying by row and BMD varying by column. d: Microscopic images of trabecular bone granules in dilute emulsion (approximately 10% FF); a number of fat droplets are visible coating the surface of the trabeculae. ................................................................................................. 153

Figure 5-2 - Delineation of oedema and fat metaplasia. Areas of bone marrow oedema and fat metaplasia were identified on T2-weighted STIR (a) and T1-weighted (d) spin echo images respectively. Freehand ROIs were placed on these images (b, e) and directly transferred to the corresponding anatomical location on PDFF maps (c, f) using the sacroiliac joints and sacral foraminae as anatomical landmarks. ................................................................. 154
Figure 5-3 - Fat–water and fat–water–bone phantoms. (a) PDFF maps from the fat–water phantom. Labels indicate reference FF values, which vary between 0 and 100%. (b) and (c) PDFF and R2* maps from the fat–water–bone phantom. Reference FF values vary by row, whereas bone mineral density values vary by column. ................................................................. 158

Figure 5-4 - Relationship of PDFF and R2* with reference FF and BMD. (a) PDFF measurements agree closely with reference FFs in the absence of bone. (b) Increases in bone density have minimal impact on PDFF measurements over a physiological BMD range. (c) R2* measurements are linearly related to BMD, although the slope varies depending on fat content. (d) Variations in FF also influence the measured R2* for a given BMD. .................. 158

Figure 5-5 - PDFF and R2* values in areas of normal bone marrow, bone marrow edema, and fat metaplasia. (a) PDFF values were significantly increased in areas of fat metaplasia, and significantly reduced in areas of bone marrow edema compared to normal subchondral marrow. (b) R2* values were significantly reduced in areas of fat metaplasia but not significantly altered in areas of edema. (c) PDFF and R2* measurements are shown on a scatterplot; areas of fat metaplasia exhibit both increases in PDFF and reductions in R2*. .............................................................................................................................................. 160

Figure 5-6 - PDFF maps (a, b) and R2* maps (c, d) demonstrating bone marrow edema and fat metaplasia in patients with sacroiliitis. Bone marrow edema (a, c; solid red arrow) causes a decrease in PDFF but no change in R2*. Fat metaplasia (b, d; dashed red arrow) causes an increase in PDFF and a reduction in R2*. The reduction in R2* may indicate a loss of bone mineral density. ................................................................................................................................. 160

Figure 5-7 - Scatterplots and Bland-Altman plots for PDFF and R2* repeated measurements. ................................. 160

Figure 6-1 - Definition of polygonal ROIs on subchondral bone. The observer is asked to define the line of the sacroiliac joint and 'anchor lines' are added to define the angles made by the joint with the anterior and posterior cortex of the bone, thus enabling the automatically-propagated ROIs to better fit the subchondral bone. ................................................................. 169

Figure 6-2 - Examples of histograms generated using the BEACH tool. Conventional MR images (a,b,c). PDFF maps (d,e,g) and PDFF histograms (g,h,i) are shown. In the normal patient's histogram (g), PDFF values are clustered around 50%, corresponding to normal marrow. In the patient with inflammation, a number of low-PDFF pixels have emerged in the histogram (h). In the patient with fat metaplasia, there is an upward shift in PDFF values, with a large number of high-PDFF pixels (i). ........................................................................................................ 171

Figure 6-3 - Examples of ADC histograms in patients with sacroiliitis (a) and control patients (b). The red lines indicate the 10th, 25th, 50th, 75th and 90th percentiles of the ADC distribution. ................................................................................................................................. 172

Figure 6-4 - ADC as an inflammatory marker ........................................................................................................ 175

Figure 6-5 - PDFF as an inflammatory marker ......................................................................................................... 176

Figure 6-6 - PDFF as a structural marker .................................................................................................................. 177

Figure 6-7 - Linear regression of qMRI parameters against visual inflammation scores. ................................. 178
Figure 6-8 - Relationship between BASDAI symptom scores and imaging measures of inflammation and structural damage (fat metaplasia). ............................................................... 179
Figure 6-9 - Geometry of ROIs propagated on subchondral bone. ........................................... 184
Figure 7-1 - Results from the fat-water-bone phantom. First-echo magnitude image and susceptibility map are shown in (a) and (b) respectively. The manually drawn circular ROIs are highlighted in red (a). Results of the 2D linear fit between bone mineral density (BMD) and fat fraction (FF) values and susceptibility are shown in (c-g). In (c), the transparent surface corresponds to measured values, while the opaque plane is the fitting 2D linear function. .................................................................................................................. 195
Figure 7-2 - Results from the fat-water phantom. First-echo magnitude image and susceptibility map are shown in (a) and (b) respectively. The linear fit between fat fraction (FF) and susceptibility values is shown in (c-d)............................................................................................................... 196
Figure 7-3 - Susceptibility maps in patients. Measured mean susceptibilities in areas of normal marrow, oedema, and fat metaplasia are shown in (a). p-values were calculated for each pair and the asterisks indicate statistical significance. Susceptibility maps and magnitude images in example subjects are shown in (b). ...................................................................................................................... 197
Figure 7-4 - Rz* maps in patients. Mean Rz* measurements in areas of normal marrow, oedema, and fat metaplasia are shown in (a). p-values were calculated for each pair and the asterisks indicate statistical significance. Rz* maps and magnitude images in example subjects are shown in (b). ...................................................................................................................... 197
Figure 7-5 - Linear regression between PDFF and susceptibility values in a single, representative subject. A rectangular ROI including fat and muscle was placed on the middle slice of the susceptibility map. Each point in the scatter plot corresponds to one voxel within this rectangular region. The blue arrow and dotted circle indicate subcutaneous fat and the corresponding points in the scatter plot. The green arrow and dotted circle indicate muscle and the corresponding points in the scatter plot. Results from the regression analysis for each subject are shown in Figure 7-7 .................................................................................................................. 198
Figure 7-6 - Nonlinear regression between PDFF and Rz* values using a quadratic model. The Rz* map and the regression of a representative subject are shown. Each point in the scatter plot corresponds to one voxel within this rectangular region. The blue arrow and dotted circle indicate subcutaneous fat and the corresponding points in the scatter plot. The green arrow and dotted circle indicate muscle and the corresponding points in the scatter plot. Results from the regression analysis for each subject are shown in Figure 7-7 .................................................................................................................. 198
Figure 7-7 - Coefficients of the linear and quadratic fits between PDFF and susceptibility (a) and Rz* (b) for each of the subjects included in the study. The fitting function is displayed in the top right corner of both subplots. In all five scatter plots, each point corresponds to one subject. Coefficients are shown as a function of the adjusted R² of each fit. The error bars indicate the 95% confidence interval of each coefficient. The slope of the linear regression (a₁) seems to be consistent in instances where the adjusted R² was high (>0.5). All other parameters had large variations across subjects. .................................................................................................................. 199
Figure 7-8 - Susceptibility measurements in patients after removing the fat contribution. 200
Figure 7-9 - R2* measurements in patients after removing the fat contribution. 200
Figure 8-1 - IDEAL-CPMG sequence as described by Janiczek et al. (209). 208
Figure 8-2 - Proposed acquisition for T2-water measurement using interleaved CPMG trains with shifted echoes. 209
Figure 8-3 - Example of segmentation using FF and R2* maps in a subject’s sacrum. The images are courtesy of Dr Jorge Cardoso, who implemented the segmentation algorithm used to generate these images. 213
Figure 10-1 - Maturity classification system. In patients with unfused intersegmental apophyses (a,b), there are persistent cartilaginous connections between the joint and neural foramina (arrowhead), and high signal bands of unossified bone adjacent to the SIJ (arrow). In the ‘partially fused’ group (c), there is partial ossification of the apophyses seen medially (arrowhead) but a thin subchondral band of high signal (arrow) remains. In the fused group (d,e), the joint margin is clearly defined (arrow) and the cartilaginous bands linking the neural foramina and SIJs have disappeared. 229
Figure 10-2 - ADC, nADC and reference ADC measurements compared by skeletal maturity. ADC measurements were significantly higher in patients with unfused sacral apophyses than those with fused apophyses (p=0.046). 232
Figure 10-3 - Study design. 241
Figure 10-4 - Examples of focal MM lesions. There is a large focal lesion in the right ischium (solid arrow) and a smaller lesion in the right ilium (dashed arrow); both lesions are shown on unenhanced Dixon images (IP, OP, WO and FO) and on the reference images (consisting of DWI and post-contrast). The smaller lesion is less conspicuous on the IP and OP images than on the FO and WO images. *DWI was acquired in the axial plane; the smaller of the two lesions (in the right ilium) is again marked with a dashed arrow. 243
Figure 10-5 - Lesion count, sensitivity, positive predictive value (PPV) and confidence for each Dixon image type. Individual observers are shown in colour (see legend), and the mean value across all four observers is shown in black. Error bars indicate the 95% confidence interval. 243
Figure 10-6 - Comparison of Percent Contrast and CNR between groups. The figures show the results of a post-hoc multiple comparison test from a one-way ANOVA. Estimates of Percent Contrast and CNR are shown as circles; the comparison intervals for each group are shown as solid lines. Percent contrast was significantly higher on FO images than on IP images (p=0.003). 248
List of Tables

Table 1 - ILAR classification of JIA................................................................................................. 50
Table 2 - ASAS criteria for positive MRI scan, according to ASAS handbook............................... 58
Table 3 - Values for chemical shift of different fat components as measured by Ren et al. (2011).
The relative shown amplitudes shown are derived from Ren et al.’s data from marrow adipose tissue, and are almost identical to those from subcutaneous fat........................................ 98
Table 4 - Median (SD) ADC, ΔADC, nADC and ΔnADC values in radiological responders and non-
responders. The two groups were compared using the Mann-Whitney-Wilcoxon test: p-
values are shown in the right-hand column. Positive ΔnADC and ΔADC values reflect an
improvement in inflammation....................................................................................................... 122
Table 5 - Median (SD) ADC, ΔADC, nADC and ΔnADC values in clinical responders and non-
responders. The two groups were compared using the Mann-Whitney-Wilcoxon test: p-
values are shown in the right-hand column. Positive ΔnADC and ΔADC values reflect an
improvement in inflammation....................................................................................................... 123
Table 6 - Adjustment for proton density of acetone......................................................................... 138
Table 7 - Results of linear regression analysis (slope and intercept values are given as estimate ±
standard error). (a) Measured PDFF values agreed closely with known fat fractions for all
BMD values in the phantom. (b) Measured R2* values were linearly related to BMD (all
slope values were significantly greater than 0), although the slope varied significantly
depending on the fat fraction. P-values relate to comparison with the 0% fat fraction vial
using ANCOVA............................................................................................................................ 157
Table 8 - Comparison of inflammatory parameters between inflamed and non-inflamed patients.
Averages are displayed as mean ± standard error of mean; AUC values are shown as value ±
standard error. The highest ROC AUC value for the evaluation of inflammation is shown in
bold.................................................................................................................................................. 176
Table 9 - Comparison of structural parameters between patients with and without fat metaplasia.
Averages are displayed as mean ± standard error of mean; AUC values are shown as value ±
standard error. The highest ROC AUC value for the evaluation of fat metaplasia is shown in
bold.................................................................................................................................................. 178
Table 10 - Interobserver variability statistics.................................................................................. 180
Table 11 – Demographics for MM patients..................................................................................... 231
Table 12 - Patient demographics, disease parameters and treatment. ISS, international staging
system; DS-PLUS, PAD, bortezomib, doxorubicin, dexamethasone; CVD, cyclophosphamide,
bortezomib, dexamethasone; VTD, bortezomib, thalidomide, dexamethasone; MPV,
melphalan, prednisolone, bortezomib........................................................................................... 238
Table 13 - Sequence parameters for WB-MRI protocol................................................................... 239
Table 14 - Results of regression analysis (n=30). Lesion count, true positives, sensitivity, positive
predictive value and confidence were compared between the four image types, using the in
phase images as the baseline. Regression analyses used image type were used as the
predictor variable, and lesion count/TP/sensitivity/confidence were used as the outcome
variable. Mean values were calculated by the regression analysis, and were equal to means
calculated manually from all patients and all four radiologists. Table 15 - Results of regression analysis for the focal lesion group alone (n=23). True positives,
sensitivity, positive predictive value and confidence are compared across the four image
types, using the in phase images as the baseline. Table 16 - Results of regression analysis for the diffuse disease group (n=6). True positives,
sensitivity, positive predictive value and confidence are compared across the four image
types, using the in phase images as the baseline.
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADC</td>
<td>Apparent diffusion coefficient</td>
</tr>
<tr>
<td>AS</td>
<td>Ankylosing Spondylitis</td>
</tr>
<tr>
<td>ASAS</td>
<td>Assessment of SpondyloArthritis international Society</td>
</tr>
<tr>
<td>BASDAI</td>
<td>Bath Ankylosing Spondylitis Disease Activity Index</td>
</tr>
<tr>
<td>BASFI</td>
<td>Bath Ankylosing Spondylitis Functional Index</td>
</tr>
<tr>
<td>CSE-MRI</td>
<td>Chemical shift-encoded MRI</td>
</tr>
<tr>
<td>DWI</td>
<td>Diffusion-weighted imaging</td>
</tr>
<tr>
<td>ERA</td>
<td>Enthesitis-related arthritis</td>
</tr>
<tr>
<td>FF</td>
<td>Fat fraction</td>
</tr>
<tr>
<td>HLA</td>
<td>Human leucocyte antigen</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>MHC</td>
<td>Major histocompatibility complex</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>PDFF</td>
<td>Proton density fat fraction</td>
</tr>
<tr>
<td>QIB</td>
<td>Quantitative imaging biomarker</td>
</tr>
<tr>
<td>qMRI</td>
<td>Quantitative magnetic resonance imaging</td>
</tr>
<tr>
<td>R2*</td>
<td>Rate constant for irreversible dephasing of spins (=1/T2*)</td>
</tr>
<tr>
<td>SEA</td>
<td>Seronegative enthesitis and arthropathy</td>
</tr>
<tr>
<td>SPARCC</td>
<td>Spondyloarthritis Research Consortium of Canada</td>
</tr>
<tr>
<td>T2*</td>
<td>Rate constant for irreversible dephasing of spins</td>
</tr>
<tr>
<td>TNFα</td>
<td>Tumour necrosis factor alpha</td>
</tr>
<tr>
<td>TNFi</td>
<td>Tumour necrosis factor inhibitor</td>
</tr>
</tbody>
</table>
1 Overview

1.1 The Clinical Need

The spondyloarthritides are a group of rheumatic diseases characterized by inflammation and new bone formation affecting the spine, large lower limb joints and entheses. Patients with spondyloarthritis suffer from a large burden of pain, morbidity and disability, which interferes with work and social function and ultimately reduces their quality of life. Both active inflammation and structural damage are thought to contribute to symptoms. Long-term outcomes are particularly poor in patients who present as children and young adults. Early diagnosis and treatment with tumour necrosis factor inhibitor (TNFi) therapy can potentially reduce the damaging effects of uncontrolled inflammation, but TNFi therapy can have harmful side effects and is expensive. It is therefore essential to accurately identify those patients with spondyloarthritis as opposed to those with other causes of pain in order to maximize therapeutic benefit and minimize side effects. Unfortunately, identifying and quantifying inflammation is relatively difficult using clinical methods alone, due to the fact that the joints are deep and cannot be palpated and to the complex, multidimensional nature of pain in this patient group. Therefore, there is an unmet need for objective methods which can accurately identify and monitor inflammation, and can be used to guide treatment decision-making in the clinic.

Magnetic resonance imaging (MRI) is commonly used for diagnosis and disease monitoring, but existing techniques are qualitative, lack reproducibility and provide only indirect information about tissue pathology. This means that scan interpretation is subjective and depends on the opinion and expertise of the local radiologist, and on the specific scanner being used. Current systems for ‘quantifying’ inflammation rely on visual assessment of inflammation using a dichotomous scoring system, which requires extensive training and calibration to achieve satisfactory reproducibility. These systems are therefore seen as impractical for clinical practice and are not widely used. The absence of an objective measure of inflammation severity can have a substantial impact on patient management, since treatment decisions are often heavily influenced by MRI findings. It is also a potential problem for clinical trials of new therapies, which currently lack objective imaging endpoints.
and are reliant on the interpretation of expert readers. Therefore, there is a need for a more objective approach to identifying and monitoring inflammation using MRI.

In this thesis, quantitative MRI (qMRI) is investigated as a candidate method.

### 1.2 Organisation of the Thesis

**Chapter 2 - Introduction**

This Chapter provides an overview of clinical and biological issues relating to spondyloarthritis, and considers how MRI could be used to address problems in clinical management pathways. The basic theory of MRI is introduced, and specific methods that are particularly relevant to the work in this thesis (particularly fat suppression, diffusion-weighted MRI and chemical shift-encoded MRI) are highlighted. The argument for using a quantitative approach to imaging inflammation in spondyloarthritis is outlined.

**Chapter 3 – The Apparent Diffusion Coefficient as a Biomarker of Inflammation**

In this Chapter, diffusion-weighted imaging is considered as a potential method for quantifying inflammation. The existing literature describing the use of ADC as an inflammatory biomarker is reviewed, and weaknesses in existing methods are considered. The development of a linear region-of-interest tool which enables an overall measure of inflammation in the sacroiliac joints is described. Subsequently, a retrospective study aiming to validate the apparent diffusion coefficient (ADC) as a marker of response to therapy in patients with spondyloarthritis is presented. Limitations of ADC as an inflammatory biomarker are also considered; some of these limitations have provided the impetus for the development of an alternative method – based on chemical shift-encoded MRI – which is discussed in the following Chapters.

**Chapter 4 – Rationale for Using CSE-MRI and Method Optimisation**

The rationale for using chemical shift-encoded MRI (CSE-MRI) to image inflammation is discussed. In order to identify specific CSE-MRI methods that can be used to for the studies subsequently presented in this thesis, a series of candidate methods are evaluated in healthy volunteers and fat-water phantoms.
Chapter 5 – PDFF and R2* as Inflammatory Biomarkers: Proof of Principle

The work presented in this Chapter describes the design of a new imaging phantom, which is used to assess the accuracy of PDFF and R2* measurements as markers of inflammation and bone mineral density. The reproducibility of PDFF and R2* measurements is assessed in healthy volunteers. In patients with spondyloarthritis, it is shown that PDFF measurements are reduced in areas of bone marrow oedema (active inflammation), and increased in areas of fat metaplasia (chronic inflammation or structural damage), compared to normal bone marrow. Additionally, the results of a phantom study and from patients suggest that R2* measurements could be used to assess bone mineral density (BMD) in trabecular bone, and could therefore play a role in quantifying new bone formation and bone destruction.

Chapter 6 – The BEACH Tool; Comparison of ADC and PDFF as Inflammatory Biomarkers in a Prospective Study

A new, partially-automated tool for measuring sacroiliac joint inflammation is proposed. This tool enables (1) automatic definition of the subchondral bone in each patient, once the observer has manually defined the joint space and (2) histographic analysis of the obtained pixel values so that active and chronic inflammation can be separately quantified. The measured parameters are validated in a prospective study of 53 patients, by comparison with radiological and biochemical measures of inflammation and clinical measures of pain and stiffness. Diagnostic performance statistics for ADC and PDFF measurements are compared.

Chapter 7 – Quantitative Susceptibility Mapping for Measurement of Bone Mineral Density

Quantitative susceptibility mapping is evaluated as a method for measuring bone mineral density, using phantom studies and data from patients with spondyloarthritis. A method for removing the contribution of fat susceptibility to the total susceptibility is described.

Chapter 8 – Discussion and Perspectives

The results of the thesis are summarized, and further work which has arisen from the work is briefly outlined. Other potential avenues for future research are discussed, and the path to clinical translation is considered.
Chapters 9 & 10 – Appendix A and B

Appendix A includes supplementary information relating to the main studies in the thesis. Appendix B includes several studies, which - for the sake of conciseness - were not been included in the main body of the thesis.

1.3 Aims and Objectives

The aim of the work presented in this thesis was to develop, implement and validate potential methods for quantifying skeletal inflammation in patients with spondyloarthritis.

Specific objectives were as follows:

1. To assess the extent to which ADC measurements reflect response to treatment in patients undergoing TNFi therapy.
2. To implement and validate a quantitative CSE-MRI protocol, which can be used to image inflamed bone.
3. To assess the accuracy of PDFF and R2* measurements as markers of fat content and bone mineral density (BMD) in a fat-water-bone phantom.
4. To compare proton density fat fraction (PDFF) and R2* measurements in inflamed bone with those in normal bone.
5. To develop a quantitative histogram-based method for measuring active and chronic inflammation in patients with spondyloarthritis, and to assess its performance in a prospective study.
6. To determine the extent to which quantitative susceptibility measurements reflect bone mineral density (and can therefore be used to monitor new bone formation) in phantoms and patients with spondyloarthritis.

Specific hypotheses relating to each of these specific objectives have been defined in the relevant Chapters; these hypotheses broadly follow previously-described frameworks for biomarker development and validation (1–3).
2 Introduction

2.1 Overview

This introductory chapter discusses the need for imaging in spondyloarthritis, and considers how MRI methods could help to meet this need.

Section 2.2 gives a broad overview of biological and clinical issues in the spondyloarthritides, and discusses potential roles for imaging in management pathways. Section 2.3 gives a basic introduction to the theory of magnetic resonance imaging techniques, with the emphasis on methods that are most relevant for this thesis (particularly fat suppression, diffusion-weighted imaging and chemical shift-encoded MRI). Section 2.4 discusses existing MRI methods for quantifying inflammation, and considers how these methods could be extended such that inflammation could be quantified.

2.2 Spondyloarthritis and the Need for Imaging

2.2.1 The Spondyloarthritides

2.2.1.1 Overview

The spondyloarthritides are a group of interrelated inflammatory disorders characterised by inflammation and new bone formation in the spine, peripheral joints and entheses (4,5). Ankylosing spondylitis (AS) is the ‘prototypic’ spondyloarthritis and causes the most dramatic phenotype, where new bone formation causes fusion of the sacroiliac, vertebral and apophyseal joints. In severe AS, fusion of these joints produces the characteristic ‘bamboo spine’ appearance on radiographs, where the whole spine becomes rigid and immobile. Psoriatic arthritis, reactive arthritis, spondyloarthritis with inflammatory bowel disease, undifferentiated spondyloarthritis and juvenile-onset spondyloarthritis also fall under the wider spondyloarthritis ‘umbrella’ (5). Vertebral inflammation (particularly inflammation of the sacroiliac joint – sacroiliitis) is a cardinal feature in all these subgroups, but fusion is variable and may be absent in some patients, particularly those with juvenile onset. The spondyloarthritides are strongly heritable, and the strongest contributing factor is the major histocompatibility complex (MHC) class I molecule HLA-B27.
Historically, classification of spondyloarthritis has been a difficult issue. Although some patients have a distinct phenotype belonging to a single subgroup (e.g. AS or enteropathic spondyloarthritis), overlapping phenotypes are common and patients may have features belonging to several subgroups. The traditional approach has been to classify patients according to their mode of clinical presentation – for example, patients with spondyloarthritis and Crohn’s disease would be labelled as enteropathic arthritis, whereas patients with spondyloarthritis and psoriasis would be labelled as psoriatic spondyloarthritis (Figure 2-1a). However, modern classification systems have aimed to use broader, simpler categories. For example, the Assessment of SpondyloArthritis International Society (ASAS) system classifies spondyloarthritis as axial (affecting the spine, pelvis and ribcage) or peripheral (affecting the extremities); the axial group is then further subdivided into radiographic and non-radiographic types depending on the appearances on plain radiography (Figure 2-1b) (6,7). There are, therefore, fewer subtypes under the modern ASAS system than in the traditional classification schemes. A potential benefit of this ‘clumping’ approach to diagnosis is that clinical treatment protocols can be simplified and applied more broadly, and expensive therapies can be licensed for a wider group of individuals. Additionally, simpler classification systems can potentially make it easier to identify patients for prognostic studies.

Figure 2-1 - Classification systems for spondyloarthritis. a. The traditional classification system emphasizes differences between subtypes. b. The modern ASAS classification system categorises spondyloarthritides into broader groups depending on axial or peripheral involvement; the axial group is further subdivided depending on the presence of ‘radiographic’ spondyloarthritis.


2.2.1.2 Epidemiology

The spondyloarthritides are predominantly diseases of young people, and approximately 80% of AS patients experience their first symptoms before the age of 30 (8). Males are affected more commonly than females, with a ratio of approximately 2 to 1 (8). AS is more common in populations where HLA B27 is prevalent, particularly in northern countries - the prevalence is highest Eskimos and Haida Indians (9). The overall prevalence of AS is thought to be between 0.1 and 1.4%, with a 1-2% prevalence for the spondyloarthritides as a whole (10,11). The incidence has been estimated to be between 0.5 and 14 per 100,000 people per year (12,13). Unfortunately, the accuracy of these estimates is limited by differences in the selection of target populations, the selection of screening criteria and the choice of diagnostic criteria between studies.

2.2.1.3 Clinical features and disease course

The characteristic symptoms of spondyloarthritis are spinal pain, stiffness and loss of mobility, particularly in the lumbar and thoracic spine, which are thought to occur due to a combination of inflammation and structural damage (the term ‘damage’ includes spinal fusion and also inflammatory sequelae such as osteoporosis) (14). However, inflammation can also occur at other locations, including the entheses (insertions of tendons or ligaments), peripheral joints and eyes, where it causes anterior uveitis. Enthesitis typically affects the plantar fascia and Achilles tendon, but can occur at multiple other locations, including the spine. Cardiopulmonary manifestations are relatively rare but include aortitis, aortic root disease, conduction disturbances and pulmonary fibrosis (15).

Spinal involvement can occur in the form of spondylitis (inflammation of the vertebra itself), spondylodiscitis (inflammation of the intervertebral disc), or spondyloarthritis itself (inflammation of the intervertebral joints). In the long term, inflammation is accompanied by structural damage due to new bone formation (osteoproliferation) which produces syndesmophytes and ultimately fusion (ankylosis) of the vertebrae. Bone loss can also occur and contributes to osteoporosis and increased fracture risk in AS (16,17). Fractures may lead to a thoracic kyphosis which is characteristically seen in male patients with advanced AS.
Peripheral joint involvement in spondyloarthritis is usually mono- or oligo-articular, with a predilection for the lower limbs, and the hip joints in particular (18). Hip arthritis is strongly predictive of a poor long-term outcome if detected during the first two years of disease (19).

Evidence regarding prognosis in AS is mixed. Several retrospective studies have suggested that spinal fusion/restriction occurs relatively early in the disease: Carette et al. showed that the 81% of AS patients having severe spinal restriction at 30-year follow up were already severely restricted when they were first assessed (20,21). However, other authors have argued that radiographic progression of disease follows a more linear course, with AS patients continuing to develop new syndesmophytes throughout the course of their disease (22). An intermediate viewpoint is that radiographic progression occurs rapidly early in the disease, and then continues at a slower pace (22).

Importantly, the rate of radiographic progression is likely to depend on disease activity (i.e. the severity or ‘amount’ of inflammation in an individual patient). Recent data from a 12-year longitudinal cohort study in 184 AS patients showed that patients with very high disease activity [as measured using the Ankylosing Spondylitis Disease Activity Index (ASDAS)], underwent substantially faster radiographic progression than those with those with inactive disease (22). Similarly, Maas et al. showed that high disease activity was associated with increased facet joint damage in the cervical spine in AS (23). Other factors associated with a severe inflammatory phenotype, such as hip disease at presentation and history of inflammatory bowel disease, uveitis or psoriasis, also increase the likelihood of radiographic progression (23,24). However, the rate of structural change does not depend solely on disease activity - unlike in other rheumatic diseases, patients without measurable disease activity do appear to structurally progress to some extent (22).

The link between inflammation and structural progression might suggest that inhibition of inflammation would prevent structural progression. However, this interpretation can be misleading (22). Multiple studies have shown that tumour necrosis factor inhibitor (TNFi) treatment has little effect on structural progression in spondyloarthritis (25–27); it is therefore unclear whether inflammation actually causes new bone formation, or whether the two processes are simply correlated.
2.2.1.4 Genetics and pathophysiology

The majority of studies examining the pathophysiology of the spondyloarthritides have focused on AS. However, many of the other subtypes share common clinical and genetic features, and it is thought that similar processes may contribute to pathogenesis in the other subgroups. The pathogenesis of spondyloarthritis is, in general, poorly understood; this subsection summarises the existing evidence and highlights the competing hypotheses regarding the cause of these diseases.

**Genetics**

There is a well-known association between AS and human leucocyte antigen B27 (HLA B27), a surface antigen involved in the presentation of peptides. One of the major functions of this protein is in antigen presentation to T-cells, which suggests that activation of the adaptive immune system might contribute to pathogenesis; however, it also possible that aberrant HLA-B27 function could indirectly activate innate immune cells (28). A potential mechanism for innate immune activation is misfolding of HLA B27 in the endoplasmic reticulum (ER), leading to an ‘unfolded protein response’ and a subsequent increase in the production of pro-inflammatory cytokines (29,30).

Despite the important link between with HLA B27, non-HLA factors (many of which remain unidentified) actually contribute the majority of the variation in overall genetic susceptibility to AS (31). Twin studies suggest that HLA B27 only contributes around 20-30% of the total genetic risk (31), whilst the risk of developing AS is much higher in HLA B27 positive relatives of AS patients than in HLA B27 positive patients without relatives for AS (32). Both these results suggest an important role for non-HLA factors in the pathogenesis of AS.

Non-HLA genetic loci associated with AS relate to diverse functions include cell activation, lymphocyte regulation, TNF signalling and peptide presentation to T-cells (4). Cytokine gene polymorphisms involving the TNF and IL-10 genes have been investigated based on their potential role in the inflammatory process, but the evidence is currently equivocal (33). Interestingly, particular cytokine polymorphisms associated with AS seem to vary depending on the population being studied – this has been attributed to the phenomenon of linkage disequilibrium,
which describes the non-random association of alleles at different loci, depending on the population (33).

**The Immune Response**

Biopsies of the sacroiliac joints in spondyloarthritis patients demonstrate dense pannus-like cellular infiltrates including T-cells (both CD4+ and CD8+) and macrophages (CD14+) with high levels of TNFα mRNA (34). The increased expression of TNFα provides a rationale for the use of TNF inhibitor (TNFi) therapies, which are commonly used to treat spondyloarthritis (see below). However, the stimulus for this aberrant immune response has not been identified.

One suggestion is that cartilage – consisting of type II collagen and proteoglycan – triggers a dysregulated adaptive immune response (34,35). In rodent models of spondyloarthritis, immunisation with proteoglycan produces spondyloarthropathy in around 70%, with a pattern similar to that seen in human AS – inflammation begins in the sacroiliac joints and then progresses to involve the intervertebral discs (36). Similarly, histological evaluation of the cartilaginous component of the sacroiliac joints of these rodents shows invasion by mononuclear inflammatory cells (37), whilst inflammation of the cartilaginous portion of the sacroiliac joints and intervertebral discs is a key feature in patients with AS (38). Furthermore, both CD4+ and CD8+ T-cells isolated from the blood and synovial fluid of patients with AS have shown responsiveness to proteoglycan aggrecan and collagen-derived peptides (39,40). Immunohistological analysis of the femoral heads of patients undergoing total hip replacement in AS demonstrates increases in T-cell, microvessel and osteoclast density in areas of subchondral bone marrow with overlying cartilage, but not at sites of complete cartilaginous destruction (41). The authors of these studies therefore argue that cartilage is a trigger for inflammation in AS, and that the bone-cartilage interface is the primary site. The reasons why cartilage might trigger inflammation in some patients, but not others, remain unknown.

**An Infectious Trigger?**

It is well known some arthritides can be triggered by infection. For example, reactive arthritis is often triggered by *Chlamydia trachomatis* infection of the genitourinary tract or by gastrointestinal *Salmonella, Shigella, Yersinia*, and *Campylobacter* infections (42). Whether there is a specific link between
spondyloarthritis and infection is less clear. The relationship between HLA-B27 positivity, spondyloarthritis and Crohn’s disease has been used as an argument for an infectious trigger by some authors (43), and in recent years there has been particular interest in the role of gut microbiota (the so-called ‘microbiome’) in both diseases (44). A potential mechanism is increased gut permeability, leading to interactions between gut microbiota and the immune system (44). There is some support for the microbial trigger hypothesis from preclinical research in animals: HLA-B27 transgenic rats develop gut and spinal inflammation resembling Crohn’s disease and spondyloarthritis respectively, but fail to develop these features when bred in a microbe-free environment (45). However, it is thought unlikely that microbial antigens would persist in typically-affected anatomical locations in spondyloarthritis, and sacroiliac joint biopsies have failed to identify candidate bacteria (46).

**ER stress**

An alternative to the adaptive immune hypothesis is that misfolding of HLA-B27 causes a stress response in the endoplasmic reticulum (‘ER stress’) which leads to immune dysregulation and increased production of IL-23 (29). IL-23 might then activate CD4+ cells expressing IL-17 (also known as T\textsubscript{H17} cells), which has been strongly implicated in several immune-mediated inflammatory diseases in animal models, including spondyloarthritis and colitis (29,30,47,48). The ER stress hypothesis is an area of current research but remains unproven at present.

**The role of the enthesis**

Although spondyloarthritis is characterised by inflammation of synovium and bone, it has recently been suggested that the entheses – the sites where tendons or ligaments insert into bone – might play a role in pathogenesis (49). The ‘enthesal stress’ hypothesis suggests that mechanical stress damages the enthesis, which in turn initiates an immune response at sites of microdamage (49). In support of this hypothesis, Jacques et al. showed that inflammation develops earliest at the entheses in a mouse model of spondyloarthritis, and that hind limb unloading suppressed inflammation compared to controls (50). Additionally, new bone formation was found to be promoted by mechanical stress at the entheses, and was correlated with the degree of inflammation (50). The argument that mechanical forces have a role in pathogenesis also receives some support from human studies: for example, Ward et al. showed that physical activities involving repetitive strain such as bending,
twisting and stretching were associated with greater functional disability and radiographic damage in AS patients (51). However, the so-called ‘enthesal stress hypothesis’ remains controversial and several unanswered questions remain. For example, it is unclear why AS is not more common in obesity and pregnancy, and why inflammation can persist in fully ankylosed spinal segments (49).

**Inflammation and new bone formation**

Although inflammation and new bone formation are regarded as central to the pathogenesis of AS, the precise nature of the relationship between these processes is complex and remains uncertain. It is unclear whether inflammation can be regarded as the cause of new bone formation, or whether the two processes occur in parallel. This issue has important implications for therapy and for drug development, since targeting inflammation alone may or may not be sufficient to prevent ongoing structural damage. One suggestion is that inflammation initially ‘triggers’ new bone formation, but new bone formation then proceeds independently via a separate pathway (52). This issue has been investigated in both preclinical and clinical research studies.

In some mouse models of spondyloarthritis, bony spur formation does clearly follow inflammation and is invariably preceded by joint erosion (53). Typically, there is a short phase of inflammation characterised by infiltration of neutrophils and mononuclear cells adjacent to the entheses, followed by a rapid cascade of endochondral bone formation (52). Enthesal cells proliferate, undergo chondrogenic differentiation and then differentiate towards osteoblasts. The subsequent process of joint ankylosis is thought to be analogous to that of embryonic endochondral bone formation, in which bone morphogenetic proteins (BMPs) have a key role. The pathological process may also be similar to fracture healing in that new bone formation occurs above periosteum near, but not at, the damaged site. In DBA/1 mice (a model of spontaneous arthritis) blockade of BMPs prevents the development of clinical arthritis, whilst active BMP signalling has also been shown in the entheses of patients with spondyloarthritis (54).

In radiological studies in adults with AS, syndesmophyte formation may not be directly preceded by inflammation, but often occurs in areas where inflammation was detected previously (26). Fat metaplasia - an increase in marrow fat content which is thought to be a sequel of inflammation - is also associated with the formation of new
syndesmophytes and with radiographic progression \((55,56)\). Nonetheless, new syndesmophytes do appear to develop at non-inflamed sites fairly commonly \((57)\). One of the limitations of studies aiming to understand the link between inflammation and new bone formation is the frequency with which patients can be scanned – it is unclear if patients develop new syndesmophytes in other areas because new bone formation is proceeding without inflammation, or because the inflammation has simply been missed because the scans were too far apart. The link between inflammation and new bone formation remains highly contentious \((58)\).

**Inflammation and bone loss**

Although the spondyloarthritides are characteristically associated with new bone formation, bone loss is also an important feature. Areas of bone loss can occur in close proximity to areas of new bone formation \((59)\), and the characteristic squaring of vertebral bodies in AS probably arises due to a combination of new bone formation and bone loss \((60)\). Bone loss is thought to be mediated by increased osteoclast activity, mediated by TNF\(\alpha\) and RANK ligand (RANKL) \((61,62)\).

Vertebral bone sections in spondyloarthritis patients show accumulations of osteoclasts with inflammatory infiltrates \((61,62)\). Unlike new bone formation, bone loss in the spondyloarthritides may be similar in pathogenesis to that observed in other inflammatory diseases such as rheumatoid arthritis, lupus and inflammatory bowel disease \((63)\). Importantly, the resulting reduction in bone mineral density is thought to contribute to the increased fracture risk in AS \((17,64)\).

**2.2.1.5 Histology**

A number of authors have sought to investigate the histological basis of spondyloarthritis, although research has been somewhat constrained by ethical limitations (sacroiliac joint biopsies are invasive and prospective studies are therefore difficult to justify). Most of this research has therefore been performed with retrospective studies including small numbers of patients.

Braun et al. performed CT-guided biopsies in five patients with AS - diagnosed using plain radiographs - and performed immunohistological analysis on the specimens \((34)\). The samples included subchondral bone and, when present, cartilage and synovium. They found dense cellular infiltrates in all five patients, which predominantly consisted of T-cells (staining positive for CD4 and CD8) and
macrophages (staining positive for CD14). The infiltrates were described as ‘pannus-like’, with islands of endochondral ossification noted between the infiltrates. It was not possible to identify normal cartilage or synovium in any of the subjects. The results of this study suggest that the normal fat-containing marrow is replaced by inflammatory cells in patients with spondyloarthritis.

Subsequently, Francois et al. performed a detailed histological study in 12 cases of AS (five biopsies and seven autopsies), who were thought to be at different stages of the disease (65). In patients who were thought to have early spondyloarthritis, inflammatory cells were found in the synovium and the subchondral marrow which was immediately adjacent (see Figure 2-2 for an example of an inflammatory infiltrate in the marrow), and small areas of pannus\(^1\) extended over the articular cartilage. In those with a longer history, the marrow became myxoid and resembled granulation tissue (defined as new connective tissue containing fibroblasts, collagen and blood vessels) (Figure 2-3). In these patients, the articular cartilage became eroded by pannus, which was in continuity with the subchondral bone marrow. The granulation tissue was seen to ‘erupt’ through gaps in the iliac cartilage in some cases, and was described as having similar appearances to the myxoid marrow (65). In patients with very longstanding disease, the joints became completely fused, and the joint space was filled with fibrovascular, scar-like granulation tissue which eroded and united the two surfaces (see Figure 2-4) (65).

Strikingly, Francois et al. also demonstrated changes in the trabecular structure of subchondral marrow in areas of inflammation. In patients with radiographic sacroiliitis, there was thinning of the subchondral end plate and trabeculae (Figure 2-5) compared to controls, while the para-articular bone (further from the joint) was sclerotic. It was suggested that these lesions might explain the erosions and sclerosis, respectively, which are observed on the radiographs of patients with spondyloarthritis (65).

\(^1\) Defined in this study as ‘an inflammatory infiltrate consisting of bone-marrow derived cells’.
Figure 2-2 - Histological image of inflamed bone marrow after staining for IL6 (a pro-inflammatory cytokine). There are a large number of inflammatory cells (arrow), with a smaller number of adipocytes (labelled fat), in the bone marrow. Reproduced with permission from (66) under License Number 4179260946789
Figure 2-3 - Histological images showing bone marrow in control (A) and AS patient (B). The iliac bone marrow (solid line) is normal in the control but has a 'myxoid' appearance in AS due to the presence of an inflammatory exudate. Reproduced with permission from (66) under License Number #179260946789.
Figure 2-4 - Invasion of the joint space by fibrovascular granulation tissue (solid arrowhead arrow, top). The cartilage (labelled C) appears to be being eroded by adjacent mononuclear inflammatory cells (open/v-shaped arrowhead, bottom). Reproduced with permission from (65) under License Number 4179280903311.

In patients with advanced AS but little disease activity, the joints were replaced with a fibrous tissue with relatively few mononuclear cells and by cartilage (containing both original articular cartilage and metaplastic chondroid tissue) (65). There was
some evidence of new endochondral bone formation. In older patients with joint ankylosis, there was often little evidence of inflammation despite complete fusion. Bony sclerosis also regressed in some of these patients.

In summary, these studies suggest that cellular inflammatory infiltrates (consisting of T-cells and macrophages) are present in subchondral bone marrow in early-stage spondyloarthritis, whilst in chronic spondyloarthritis the marrow is myxoid and contains fibroblasts. The joint space is gradually filled with granulation tissue (pannus), which contributes to joint erosion and eventual fusion.

2.2.1.6 Spondyloarthritis in children and adolescents

Classification of paediatric spondyloarthritis

Although it is widely acknowledged that the spondylarthritides form a continuum that encompasses a range of interrelated inflammatory diseases, patients in a paediatric age range (which typically extends up to 16-18 years) have historically been described and classified separately to adult patients. Arguably, this has led to patients in this age group being relatively neglected, and has hampered research efforts and the development of clinical guidelines, particularly regarding the use of biologic therapy (67). Young patients with spondyloarthritis have been variously classified as ‘juvenile spondyloarthritis’, ‘juvenile spondyloarthropathy’ and ‘juvenile ankylosing spondylitis’ without accepted diagnostic criteria² (67).

These classification systems have been suboptimal for patients with childhood-onset spondyloarthritis because the clinical features are somewhat different to those in adults (67). Peripheral involvement, including enthesitis and peripheral arthritis, are common features of childhood-onset spondyloarthritis but axial symptoms are relatively rare, at least until later in the disease course (67). This means that childhood-onset spondyloarthritis can be confused with other forms of juvenile arthritis and therefore treated inappropriately.

In recognition of differences between adult and juvenile spondyloarthritis, Rosenberg and Petty developed a set of criteria required for diagnosis of

² In this thesis, paediatric patients with spondyloarthritis are referred to using the term ‘childhood-onset spondyloarthritis’ – which is defined here as spondyloarthritis presenting before the age of 16 - for the avoidance of ambiguity.
'seronegative enthesopathy and arthropathy' (SEA), which included the presence of enthesitis with arthralgia or arthritis, the absence of rheumatoid factor and antinuclear antibodies, and the onset of symptoms before 17 years of age (68). This category recognized that childhood-onset spondyloarthritis could occur in the absence of axial symptoms, and was therefore useful in differentiating this group of patients from those with juvenile rheumatoid arthritis. More recently, the ILAR criteria aimed to identify relatively homogenous sets of patients with juvenile arthritis which could be used for research purposes (Table 1); under this system, young patients with spondyloarthritis fall into the enthesitis-related arthritis (ERA) category (69). The definition of ERA is similar to that of SEA, but with some distinctions – for example, arthritis and enthesitis together is sufficient for the diagnosis of ERA (but not SEA), and ERA has an upper age limit of 16. ERA reportedly accounts for approximately 10-20% of patients with JIA (70,71), although this number is typically higher in adolescent cohorts (67).

Whilst the ILAR system is an improvement over previous classification schemes, it still has limitations. There is no specific classification for children who meet the criteria for AS, even though a high proportion of AS patients experience symptoms before the age of 16 (72). Additionally, patients with psoriatic spondyloarthritis are not clearly differentiated from patients with psoriasis and peripheral joint involvement. Finally, although ERA has historically been viewed as a peripheral enthesitis at disease onset, several recent studies have demonstrated the occurrence of axial inflammation relatively early in the disease course. Pagnini et al. showed that 30% of patients already had evidence of sacroiliitis on MRI at a median of 1y3m after disease onset, suggesting that axial disease may often be missed (73). The prevalence of axial involvement may depend on the patient’s age – in adolescents, up to 30% of patients may have inflammatory spinal symptoms at disease onset (74). A commonly-held view is that there are two distinct clinical phenotypes in ERA – those with early axial disease (often associated with hip arthritis) and those with predominantly peripheral disease who do not develop axial disease or develop it after many years (74). However, further research is required to clarify this distinction, and to explain why the mode of presentation is different for these groups in terms of pathogenesis and/or genetics.
<table>
<thead>
<tr>
<th>Subtype</th>
<th>Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systemic arthritis</td>
<td>Arthritis in more than one joint, and fever for at least two weeks with one or more of the following: rash, lymphadenopathy, hepatosplenomegaly, serositis</td>
</tr>
<tr>
<td>Persistent oligoarthritis</td>
<td>Arthritis affecting four or fewer joints during the first six months of disease, who do not develop arthritis in additional involved joints over time</td>
</tr>
<tr>
<td>Extended oligoarthritis</td>
<td>Arthritis affecting four or fewer joints during the first six months of disease, who subsequently experience involvement of more than four joints</td>
</tr>
<tr>
<td>Rheumatoid-factor-positive</td>
<td>Arthritis affecting five or more joints during the first six months of disease, rheumatoid factor positive</td>
</tr>
<tr>
<td>Rheumatoid-factor-negative</td>
<td>Arthritis affecting five or more joints during the first six months of disease, rheumatoid factor negative</td>
</tr>
<tr>
<td>Enthesitis-related arthritis</td>
<td>Arthritis and enthesitis</td>
</tr>
<tr>
<td></td>
<td>OR</td>
</tr>
<tr>
<td></td>
<td>Arthritis or enthesitis with two of the following:</td>
</tr>
<tr>
<td></td>
<td>Sacroiliac joint tenderness, inflammatory spinal pain or both</td>
</tr>
<tr>
<td></td>
<td>HLA B27 positivity</td>
</tr>
<tr>
<td></td>
<td>Family history in 1st degree relative of medically confirmed HLA-B27 associated disease</td>
</tr>
<tr>
<td></td>
<td>Acute anterior uveitis</td>
</tr>
<tr>
<td></td>
<td>Onset of arthritis in a boy after the age of 6 years</td>
</tr>
<tr>
<td>Psoriatic arthritis</td>
<td>Arthritis with psoriasis, or arthritis with at least two of the following:</td>
</tr>
<tr>
<td></td>
<td>Dactylitis</td>
</tr>
<tr>
<td></td>
<td>Nail pitting/onycholysis</td>
</tr>
<tr>
<td></td>
<td>Psoriasis in a first degree relative</td>
</tr>
</tbody>
</table>

Table 1 - ILAR classification of JIA.
A common pathogenesis?

Despite the differences in the mode of clinical presentation, it has been argued that the aetiology of juvenile spondyloarthritis is closely related to that of adult spondyloarthritis (75). Between 76 and 85% of patients with ERA are HLA B27 positive, suggesting a common pathogenetic mechanism with adult AS (76,77). However, some subtypes of HLA B27 (e.g. B2705) are more common in children with spondyloarthritis, and it has been argued that subtle differences between genotypes could account for the earlier presentation in patients with childhood-onset spondyloarthritis (78). In support of this idea, AS and ERA are associated with single nucleotide polymorphisms (SNPs) in the ERAP1 gene, which encodes an endoplasmic reticulum aminopeptidase involved in trimming and optimizing peptides for presentation by MHC I molecules, and in the cleavage of pro-inflammatory cytokines (79). ERAP1 polymorphisms might cause disease by altering the peptide repertoire of HLA-B27, thereby altering the presentation of arthritogenic peptides by immune cells (29). Similarly, the psoriatic form of JIA shares a link with IL23 receptor SNPs with adult psoriatic arthritis (29).

2.2.1.7 Relationship between age at onset and prognosis

There is some evidence that the prognosis of patients with spondyloarthritis varies depending on the age of onset. A study of 546 patients with childhood-onset AS (≤16 years), adult-onset AS and late-onset AS found that patients with childhood onset had higher serum immunoglobulin levels (suggesting more severe inflammation) and poorer clinical outcomes (80). Juvenile-onset patients are also more likely to require surgery in later years than patients with adult-onset spondyloarthritis (81,82). However, other studies have found that adult-onset disease is associated with worse functional and quality of life measures once disease duration is accounted for (83). Gensler et al. did not identify a significant difference in function between adult-onset and childhood-onset groups, but found that childhood-onset patients tended to have more severe hip involvement and less severe axial involvement than adult-onset patients (84).

Amongst JIA subtypes, patients with childhood-onset spondyloarthritis (ERA) are amongst those with the poorest prognosis and very often require treatment into adult life (81,85). In a study from Berlin in 2002, around 20% of ERA patients achieved treatment-free remission at long-term follow up (median follow up time
16.5 years) (86). This difficulty with withdrawing treatment has also been reported in a randomized controlled trial from Germany, which shows that ERA patients who respond to etanercept are significantly more likely to experience a disease flare if the treatment is withdrawn compared to when it is continued (87).

**A chance for early intervention?**

Although the development of TNFi agents has dramatically improved the outlook for patients with spondyloarthritis in terms of pain, stiffness and disability, several studies have found that TNFi treatment does not slow the rate of radiographic progression in patients with established spinal ankylosis (25,27). This has led to a drive to identify patients with ‘pre-radiographic’ spondyloarthritis, in whom it might be possible to halt or modify the disease early and therefore prevent late-stage disease. Spondyloarthritis in children is, by definition, early-onset and therefore arguably provides an opportunity to intervene early in the disease course, before the development of structural damage (67). However, further studies are required to support this hypothesis.

### 2.2.1.8 Treatment of spondyloarthritis

A detailed review of treatment strategies in spondyloarthritis is beyond the scope of this work, but a brief summary is given here.

In axial spondyloarthritis, the major treatment goals are to reduce pain, improve and maintain spinal flexibility and posture, minimize functional limitations and minimize the impact of the disease on quality of life. Management guidelines have been issued by consensus committees in Europe, the United States, and Canada (88–90). Although there are some subtle differences between these guidelines, the general principles of treatment are widely agreed upon.

Patients are advised to follow an active exercise program, designed to maintain posture and improve mobility, irrespective of whether their disease is active or stable.

The first line drug therapy is with non-steroidal anti-inflammatory drugs (NSAIDs). NSAIDs are inhibitors of the enzyme cyclooxygenase (COX), and therefore reduce the production of prostaglandin and thromboxane from arachidonic acid. Continuous treatment with NSAIDs is recommended for patients with persistently active and
symptomatic disease, whereas 'on demand' treatment is typically used for patients with stable spondyloarthritis, or when continuous treatment causes unacceptable side effects. There is little evidence for any particular NSAID over another in terms of efficacy, although side-effect profiles may vary between drugs.

In adult spondyloarthritis patients whose disease is poorly controlled, TNFi therapy is recommended. The TNFi category encompasses a group of drugs that are designed to suppress the physiological response to TNF, which is a pro-inflammatory cytokine. The majority of the TNFi group (infliximab, adalimumab, certolizumab and golimumab) are monoclonal antibodies which bind to the TNFα receptor and thereby inhibit the effective binding to TNFα. Additionally, this group also includes a circulating receptor fusion protein (etanercept) which acts as a 'decoy' receptor for TNF and therefore reduces the effect of the naturally-present TNF in the tissues. There are numerous randomized controlled trials now showing that TNFi therapy produces rapid and sustained improvement in disease activity and function, which have been reviewed in Ref. (91). Around 60% of patients are thought to have an 'adequate and usually sustained' response to TNFi, although a minority experience a smaller benefit or do not respond to therapy (91). Response to treatment remains somewhat unpredictable, but a short disease duration, early age at onset, raised inflammatory markers and good baseline function are thought to increase the probability of response. Response to TNFi has not been found to differ

**Figure 2-6** - TNF inhibitors. Fv - variable region, Fc - fragment crystallisable region (interacts with immune system).
significantly between non-radiographic axial spondyloarthritis and AS, although regulatory approval for these indications has varied between countries (91,92).

Patients with juvenile-onset spondyloarthritis are managed slightly differently to adult patients. Whereas disease modifying anti-rheumatic drugs (DMARDs) such as methotrexate are rarely used for isolated axial disease in adults, patients with juvenile-onset disease are often treated with DMARDs prior to starting TNFi therapy. This difference is partly due to the classification of juvenile-onset spondyloarthritis as a subtype of JIA (i.e. the ERA subtype), since methotrexate is used commonly as a treatment for JIA in general, and partly due to drug licensing constraints which are specific to the paediatric age group (67). Anecdotally, methotrexate does not seem to be as efficacious as TNFi therapy in this group and is often poorly tolerated, with frequent reports of nausea and behavioural complaints (93). However, there are no good studies evaluating the efficacy of methotrexate in juvenile-onset spondyloarthritis/ERA. Response to methotrexate may be partly determined by an individual’s specific inflammatory phenotype, and several authors have sought to develop biomarkers which will predict methotrexate response (issues around response prediction have been discussed further in Section 2.2.2.3) (94). None of these markers are being used clinically at present.

In any age group, patients with local inflammation at a specific site (for example, the sacroiliac joints) can be treated with local glucocorticoid injection. The systemic use of glucocorticoids is generally avoided because of the increased risk of vertebral osteoporosis, which may add to the fracture risk which is already associated with spondyloarthritis. Sulfasalazine may be useful for patients with spondyloarthritis who have peripheral involvement.

2.2.1.9 Serum biomarkers in spondyloarthritis

A biomarker is ‘a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic process, or pharmacologic responses to therapeutic intervention’ (95). In spondyloarthritis, there have been efforts to develop biomarkers which can measure disease activity and can therefore predict joint damage or response to treatment (which is expensive and has potentially harmful side effects) (96). Unfortunately, common biochemical markers such as C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR) lack sensitivity and specificity in spondyloarthritis (96). Clinical parameters such as age
and baseline functional status do carry some predictive information, but this is limited (97). Therefore, a number of soluble biomarkers - detected in peripheral blood and urine - have been investigated.

A detailed review of these soluble biomarkers is beyond the scope of this thesis, and can be found elsewhere (96). Briefly, candidate markers include high-sensitivity CRP, which has been shown to correlate better with symptoms than routine CRP (98), and other inflammatory cytokines such as TNFα and interleukins 6, 7, 17 and 23 (96). Matrix metalloproteinase-3 and sclerostin have shown some potential for predicting damage (99,100). Dickkopf-1 (DKK-1) is involved in regulating new bone formation and has been extensively investigated in spondyloarthritis, but the evidence has been inconsistent and DKK-1 levels have variously been reported to be elevated, reduced or unchanged in spondyloarthritis patients compared to controls (101–103).

In general, research into these serum biomarkers is at an early stage; most studies have used only small sample sizes and have not been able to show that the biomarkers contribute information that is independent of clinical or laboratory variables used in standard care. None of these biomarkers are routinely used in clinical practice. Nonetheless, the large number of studies that have been performed in this area does highlight that there is an unmet need for noninvasive methods for measuring inflammation and structural damage in spondyloarthritis.

2.2.2 Why Image? – and the Case for MRI

A central theme of this thesis is the use of imaging to improve the management of patients with spondyloarthritis. In this Section, specific clinical situations in which imaging may be helpful are highlighted. Limitations of existing approaches are discussed, and particular attention is given to areas where imaging techniques or approaches could be improved.

It should be noted that the majority of research into the use of MRI in spondyloarthritis comes from adults with AS or axial non-radiographic spondyloarthritis. Although much of this research can be extrapolated to children, there has been far less research in this patient group.
2.2.2.1 Diagnosis

One of the central difficulties with managing spondyloarthritis is that diagnosis is difficult; as a result, a long delay between symptom onset and diagnosis is the norm. In a study by Feldtkeller et al., the average time between symptom onset and diagnosis was 8.5 years for HLA B27 positive patients, and 11.4 years for HLA B27 negative patients (8). Spondyloarthritis can be particularly difficult to diagnose in young patients, due to differences in the mode of presentation (peripheral features may predominate), the nonspecific nature of back pain in this group and the fact that the sacroiliac joints are deep and difficult to palpate (67). Clinical assessment of sacroiliitis specifically is surprisingly unreliable compared to imaging (104,105) and blood biochemical markers may be normal in active disease (106,107). An imaging test which can diagnose sacroiliitis with high sensitivity and specificity is therefore needed.

Plain radiography has been used to diagnose ankylosis of the sacroiliac joints and spine since the 1930s, and is still recommended as a first line imaging method to diagnose spondyloarthritis in adults and those patients with longer symptom duration (108). However, radiography can only detect structural changes such as ankylosis and sclerosis, which are indicators of damage rather than inflammation per se and can take a number of years to develop (108). Even for these structural changes, the reliability of visual assessment between observers is limited, with typical kappa statistics in the range of 0.16-0.56 for syndesmophytes and 0.48-0.95 for ankylosis (109). As a result, plain radiography is prone to generating false positive results (110), and reported statistics for the sensitivity and specificity of radiography are inferior to those for MRI and CT (111,112). Therefore, in recent years there has been a move away from using plain radiography in spondyloarthritis, particularly in young patients who have early disease where inflammation – rather than structural damage – is the predominant feature (67).

Some authors have used ultrasound to detect sacroiliitis, although there are a number of technical limitations with this approach - in particular, the sacroiliac joints are deep and the adjacent bone is highly echogenic, making assessment difficult (104,113). Ultrasound can be used to assess peripheral involvement in spondyloarthritis – including peripheral arthritis, tenosynovitis and enthesitis – but operator expertise can be variable and interpretation relies on subjective
identification of particular features, depending on the clinical context (113).
Furthermore, recent data from the Devenir des Spondyloarthrites Indifférenciées Récentes (DESIR) cohort of 708 patients with spondyloarthritis did not find a significant association between ultrasound assessments of disease activity and clinical or biochemical measures of inflammation, arguing against the use of this tool in clinical practice (114).

MRI is well-suited to measuring inflammation since it is intrinsically sensitive to tissue water content. In areas of active inflammation, there is an increase in water content and a reduction in the fat content of bone marrow, since normal fatty marrow is replaced by water and cells (38). MRI can also detect chronic inflammatory/structural changes such as fat metaplasia (defined as an increase in subchondral marrow fat content relative to normal marrow), which is thought to represent post-inflammatory damage (115). MRI is significantly more sensitive than plain radiography for detecting sacroiliitis as determined using clinical criteria (116). Furthermore, inflammatory changes on MRI can precede the development of structural changes, suggesting that MRI may be helpful for the detection of 'early' disease (117). MRI-diagnosed sacroiliitis is associated with longer disease duration, higher inflammatory markers and more severe back pain (118).

MRI is now recommended as the first-line imaging modality for the diagnosis of sacroiliitis in adults by ASAS, who also recommend a set of MR sequences for optimal detection of inflammation, and specific imaging features which the radiologist should look for (see also Section 2.4 for a discussion of the biologic basis of image contrast in skeletal inflammation) (119). A number of studies have now been published demonstrating the diagnostic accuracy of MRI for the diagnosis of spondyloarthritis in adults (120–124), although evidence is lacking in paediatric and adolescent patients.

Unfortunately, a problem with many of the studies investigating the diagnostic accuracy of MRI in spondyloarthritis is that the MRI scan contributes to both the reference standard and the index test (125,126). A further potential source of bias is that some of the investigators performing these studies have a pre-determined view regarding the inflammatory changes on MRI which are suggestive of spondyloarthritis (particularly as many of them were involved in the development of the criteria for the positive MRI scan) (117,127). In populations which are not pre-
selected - i.e. the patients have back pain but not necessarily sacroiliitis - the association between MRI findings and clinical features is much less clear than in cohorts where patients are known to have spondyloarthritis (128). Undoubtedly, there is a need for more high-quality studies exploring the relationship between MRI findings and clinical features in spondyloarthritis.

A specific problem with using MRI for diagnosis is that the definition of a ‘positive’ MRI (defined as scan which is ‘highly suggestive of spondyloarthritis’) is rather vague. ASAS suggest that bone marrow oedema must be ‘highly suggestive of spondyloarthritis’ and ‘clearly present’ without further explanation, meaning that there is considerable scope for interpretation of these recommendations (119). Similarly, it is suggested the inflammatory ‘lesions’ should be present on at least two consecutive slices (or should be multiple on a single slice) (119), but it is unclear how these recommendations relate to lesions which are confluent, and how the interpretation should be modified depending on the acquisition slice thickness. This problem has been highlighted by Deodhar, who commented that “a tiny bit of white on two consecutive slices’ may be objective, but not specific” (127). The rather vague definition of a positive MRI scan has also been a problem for drug licensing bodies such as the US Food and Drug Administration and the European Medicines Agency (129).

The lack of objectivity surrounding the definition of a positive MRI scan is one reason for developing quantitative methods for measuring inflammation; this point has been discussed further in Section 2.2.3.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bone marrow oedema</td>
<td>‘The presence of definite subchondral bone marrow oedema/osteitis highly suggestive of sacroiliitis is mandatory’</td>
</tr>
<tr>
<td>Synovitis, capsulitis or enthesitis</td>
<td>‘The presence of these features alone, without concomitant subchondral bone marrow oedema, is compatible with sacroiliitis but not sufficient for making a diagnosis of active sacroiliitis’</td>
</tr>
<tr>
<td>Technical aspects</td>
<td>‘STIR images are usually sufficient to detect active (acute) inflammatory lesions; exception: synovitis (not detectable with STIR only, T1 post-gadolinium is needed)’</td>
</tr>
<tr>
<td>Amount of signal required</td>
<td>‘If there is one signal (lesion) only, this should be present on at least two slices. If there is more than one signal on a single slice, one slice may be enough’</td>
</tr>
</tbody>
</table>

Table 2 - ASAS criteria for positive MRI scan, according to ASAS handbook.
2.2.2.2 Measuring disease activity

The aim of treatment in spondyloarthritis is to minimize inflammation in the sacroiliac joints and spine, in order to reduce pain and prevent long term damage. A number of clinical questionnaires have been designed to assess disease activity and function, including the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), Bath Ankylosing Spondylitis Functional Index (BASFI) and Ankylosing Spondylitis Disease Activity Score (ASDAS) (130–132). However, quantifying disease activity using clinical assessment is problematic because inflammatory and biomechanical back pain may co-exist in the same patient, meaning that patients with predominantly biomechanical back pain can have high clinical scores. Additionally, clinical symptoms are often vague and poorly localized. The poor specificity of clinical disease activity measures may be particularly pronounced in childhood-onset groups because of the complex, multidimensional nature of pain in young people (133). Pain may be purely ‘inflammatory’ in some patients but anatomical, psychosocial, environmental and social factors may be more important in others (133). This makes targeting therapy very difficult – some patients may benefit from early and aggressive anti-inflammatory therapy, whilst others may benefit more from pain management strategies including pharmacological, physical (including sleep) and psychological therapies (133).

A number of studies have demonstrated the utility of MRI for measuring disease activity, and the European League Against Rheumatism (EULAR) now recommend that MRI can be used for this purpose (108,134,135). In longitudinal studies, changes in inflammation have been reported as early as 6 or 12 weeks after biologic therapy (134,136). There is some debate about whether to use contrast-enhanced imaging for disease monitoring; current recommendations suggest that short tau inversion recovery (STIR) imaging alone is sufficient (see Section 2.3 for a detailed discussion of imaging sequences, including STIR imaging) (108).

A further potential benefit of using MRI for disease monitoring is that it may improve patient satisfaction and/or compliance. There is increasing awareness that young people with arthritis have a desire for more disease-specific knowledge, and a more ‘individualised’ approach to therapy may therefore improve psychological outcomes (133,137).
2.2.2.3 Prediction of therapeutic response

It might be expected that patients with higher levels of inflammation would respond better to biologic therapy, and MRI has therefore been investigated as a tool for predicting clinical response. Rudwaleit et al. found that MRI spinal inflammation scores were predictive of achieving a 50% reduction in BASDAI score in patients participating in randomized controlled trials (138). Similarly, Sieper et al. found that patients in the ABILITY-1 trial of adalimumab were more likely to respond if their baseline inflammation scores were ≥2 (139). However, Barkham et al. found no difference in clinical response in patients between patients with mild and moderate/severe sacroiliitis (140).

Interestingly, the extent to which an individual responds to treatment may depend both on that individual’s genotype and/or phenotype and on the treatment in question. Specific single nucleotide polymorphisms and pro-inflammatory cytokines (such as the pro-inflammatory S100 proteins) have both been shown to influence the likelihood to respond to methotrexate, although at present none of these serum or genetic markers have reached clinical use (94). It is possible that combining clinical, genetic and imaging data into a predictive model for therapeutic response could help to identify those who would benefit most from specific treatments.

2.2.2.4 Prediction of sustained remission

A particular problem with young spondyloarthritis patients is that sustained remission is difficult to achieve. Ideally, it would be possible to withdraw treatment once inflammation is adequately controlled, but in practice inflammation often recurs after treatment withdrawal (86). Between 30-50% of JIA patients treated with methotrexate relapse upon treatment withdrawal, and this percentage is likely to be higher in childhood-onset spondyloarthritis/ERA patients (141,142). Several authors have investigated whether inflammatory biomarkers could be used to predict stable remission, particularly in the context of methotrexate therapy (94). In general, it seems that patients with higher levels of inflammation at the time of cessation are more likely to relapse, although more research is required to fully test this hypothesis (94). The main focus of these studies has been serum biomarkers, but imaging biomarkers could potentially be useful given their sensitivity and specificity for relatively low levels of inflammation.
2.2.2.5 Prediction of phenotype and severity

An important aim of therapy in spondyloarthritis is to prevent structural damage in the form of ankylosis and bone destruction. However, at presentation it may be difficult to identify those patients who will go on to ankylose as compared to those who will have a relatively indolent disease course.

Baseline radiographic changes (particularly syndesmophytes) do appear to be a strong predictor of radiographic progression in AS (108,143–145). Likewise, MRI studies have demonstrated correlations between vertebral corner inflammatory lesions and subsequent syndesmophyte formation in the spine (146,147), and between baseline SIJ inflammation or fat deposition with radiographic progression in the SIJs (148). The results of these imaging studies are consistent with those of studies measuring disease activity clinically, where there is also a correlation between disease activity and the severity of structural damage in the spine in AS (149). In standard clinical practice, assessments of disease activity do contribute to decision making in a subjective fashion but there is currently no objective mechanism for stratifying treatment according to severity.

2.2.2.6 Monitoring structural changes

**Ankylosis**

Several authors have shown that structural damage (particularly ankylosis) can be monitored using plain radiographs (108,150–152). There are fewer studies examining the use of MRI for detecting structural damage, although MRI features correlate with those observed on radiography or CT (153). MRI scoring systems for structural damage have been developed by Madsen et al. and Maksymowych et al., and include features such as erosions, fat metaplasia (an increase in fat content on T1-weighted images), and joint fusion (154,155). However, a problem with these methods is that scoring is dependent on the expertise and opinion of the observers, and inter-rater agreement is relatively poor (155). Additionally, the MRI methods used in these studies do not enable measurement of bone mineral density, and therefore lack much of the structural information provided by CT scans.

**Osteoporosis and spinal fractures**

Patients with spondyloarthritis may develop osteoporosis due to systemic effects of inflammation and/or steroid use, which contributes to the increase fracture risk seen
in AS (17,156). Bone mineral density (BMD) is commonly assessed using dual energy X-ray absorptiometry (DXA) (108), although quantitative CT is seen as a ‘gold standard’ and detects more cases of osteoporosis than DXA in patients with advanced AS (157,158). However, neither of these techniques are used frequently in clinical practice, due to their relative inconvenience and the need for ionizing radiation. Recent proof-of-principle studies suggest that MRI could also be used to detect osteoporosis using quantitative imaging biomarkers such as R2* (159); potential imaging biomarker of bone formation in spondyloarthritis have also been investigated in this thesis.

2.2.2.7 Limitations of MRI

Although MRI has a number of significant advantages compared to other modalities, it does have limitations. Acquisition times are relatively long compared to other modalities (the spondyloarthritis protocol takes around 50 minutes at our institution), which can be particularly problematic for patients with back pain. Additionally, administration of contrast medium requires the insertion of an intravenous cannula, which can be painful and is a significant cause of anxiety in young patients (160,161). Claustrophobia is a problem and can lead to premature termination of scans, which may be distressing for the patient and lead to incomplete diagnostic information (162). The scanners themselves are also expensive to purchase and maintain. Overall, no imaging modality is without its flaws and the benefits of MRI are generally perceived to outweigh the disadvantages, leading to widespread acceptance of this technique as part of clinical pathways (108). One particular limitation of MRI which could potentially be improved is the qualitative nature of the scans, which is discussed in the following section.

2.2.3 The Argument For A Quantitative Imaging Biomarker

2.2.3.1 Problems with qualitative image assessment

The vast majority of published studies using imaging in spondyloarthritis have relied on qualitative assessment of images – i.e. the scans are assessed by a radiologist or rheumatologist, who ‘scores’ the images based on their impression of whether a particular feature is present on a particular slice or stack of images. However, this approach suffers from a number of problems.
A widely-used example of the qualitative approach is the Spondyloarthritis Research Consortium of Canada (SPARCC) system, which assesses the severity of inflammation in the sacroiliac joints (163). The SPARCC method relies on subjective assessment of STIR images by a radiologist, who makes a binary decision as to whether inflammation is present in each joint quadrant, on each slice. If inflammation is present, a score of 1 per quadrant is assigned; if there is no inflammation, a score of 0 is assigned. Additional 'points' are added if patches of inflammation are particularly intense (i.e. the signal intensity is greater than or equal to that of presacral blood vessels) or measure more than 1cm in depth from the joint. A labelled MR image highlighting key features of the SPARCC method is shown in Figure 2-7.

![SPARCC Scoring System](image)

Figure 2-7 - SPARCC scoring system. Each sacroiliac joint is divided into four quadrants, and the observer allocates a point for each quadrant where inflammation is present. Extra points are given if the inflammation (solid arrow) is at least as intense as presacral blood vessels (dashed arrow), or if the signal changes extend more than 1cm from the joint surface.

Unfortunately, this visual scoring system introduces a large subjective element and means that the assigned score is dependent on the expertise (and opinion) of the radiologist. The binary choice for each joint quadrant is particularly unsatisfactory where only early/subtle inflammatory changes are present. Image interpretation also depends on the quality of fat suppression, which is variable and depends on the specific sequence and scanner being used. MR imaging is susceptible to artifacts, and relatively small artifacts can be interpreted as inflammation, particularly by observers who are inexperienced or are unfamiliar to the specific scanner being used.
These issues can lead to poor agreement both within and between observers, even in controlled research settings where observers have often undergone calibration exercises prior to participation.

One of the most comprehensive studies in this area was undertaken by Lukas et al., who performed a multi-reader study examining the inter-reader reliability, sensitivity to change and discriminatory ability of 3 different scoring methods (164). Using 9 experienced readers, they found inter-reader intraclass correlation coefficient (ICC) values of 0.49 - 0.77 for disease activity status and 0.46 - 0.72 for disease activity change (164). However, when specific reader combinations were evaluated, ICC values varied dramatically and ranged from 0.05 (very low agreement) to 0.90 (excellent agreement), indicating a high degree of operator-dependence. This operator-dependence was present despite the use of a detailed training session prior to starting the study, and the fact that majority of the readers in the study were described as spondyloarthritis experts.

Similarly, Laloo et al. recently reported kappa statistics of 0.64, 0.65, 0.81 and 0.52 for inter-observer agreement for high T1 signal, fluid signal, ankylosis and vacuum phenomena assessed on MRI (121). Arnbak et al. found kappa values ranging from 0.36 to 0.81 for a range of features in the spine and sacroiliac joints (165). In this study, the kappa statistic for overall diagnosis of spondyloarthritis was 0.61 (165). This figure is at the lower end of the ‘substantial agreement’ category using conventional description systems for Kappa statistics, although authors such as McHugh et al. have argued that this categorization system can be too lenient in healthcare research; under a proposed alternative scheme the value of 0.61 would be described as moderate agreement (166).

In clinical practice, images are evaluated by a variety of observers who may be musculoskeletal radiologists, non-specialist radiologists or rheumatologists. The ‘score’ provided by these observers depends on the expertise and opinion of the observers, their clinical specialty and the health service in which they work. Additionally, perceptions of inflammation severity may be biased by the nature of the caseload at a given hospital – clinicians who are used to seeing many ‘severe’ cases of spondyloarthritis may give systematically lower ratings than those who only see the disease rarely. These factors make qualitative scoring systems rather unsatisfactory for use in clinical practice, especially when treatment decisions...
regarding expensive and potential harmful therapies are involved. The lack of objectivity is also a problem for research, because studies aiming to evaluate the predictive or prognostic capability of a particular image feature or diagnosis may be biased by the observers who participate in that study.

2.2.3.2 Quantitative imaging biomarkers

In the last decade, there has been a rapid expansion in the use of quantitative MRI techniques to measure disease characteristics (1,167–170). These techniques typically use a succession of scans to infer tissue attributes such as cellularity, vascularity or fat content, based on the change in signal characteristics over these scans. Each pixel (picture element) in a parametric map has a measurable numerical value that reflects the intrinsic properties of a tissue, rather than an arbitrary signal intensity produced by standard MRI. In this way, quantitative MR images can be viewed as a set of measurements which are analogous to measurements made by laboratory assays. These measurements can be used as quantitative imaging biomarkers. Sullivan et al. define a quantitative imaging biomarker (QIB) as, 'an objectively measured characteristic derived from an in vivo image as an indicator of normal biological processes, pathogenic processes, or response to therapeutic intervention' (1).

QIBs have a number of advantages over qualitative scoring systems such as the SPARCC method. Most obviously, using QIBs can eliminate the need for visual interpretation, thus avoiding the bias, inconsistency and operator-dependence which are inherent to qualitative scoring. Additionally, QIB measurements can be automated relatively simply, and could potentially be incorporated into scanner software or made available on PACS workstations. This would mean that objective measurements of disease characteristics would be available to clinicians much more readily and might lower the threshold for introduction into clinical practice.

2.2.3.3 New insights into tissue characteristics

A further potential advantage of using quantitative imaging is that it may reveal some tissue properties not reflected using conventional imaging. Whereas contrast in 'conventional' MR images is influenced by a variety of tissue properties (including the proton density, relaxation times T1 and T2 and the freedom of water diffusion – see Section 2.4, below), quantitative MRI can theoretically assess each of these
properties in isolation. For example, diffusion-weighted imaging (DWI) produces apparent diffusion coefficient (ADC) maps which reflect the freedom with which water molecules diffuse, whilst dynamic contrast-enhanced MRI can give objective estimates of tissue perfusion, whilst T2 and T1 mapping can actually estimate the values of these relaxation constants in the tissue. Imaging biomarkers which are more specific for underlying processes may enable us to gain a better understanding of pathophysiology, and to treat individual patients more effectively. In the case of spondyloarthritis, there is a need for an imaging method which reflects bone density, and quantitative methods such as R2* or susceptibility mapping may be able to assess this property.

Despite the potential advantages of QIBs, there are some potential disadvantages to using these methods, which have been discussed in detail in the relevant Chapters of this thesis (particularly Chapters 2, 4 and 5).

2.3 Introduction to Magnetic Resonance Imaging

2.3.1 Introduction

Having described the clinical features of spondyloarthritis and the unmet need for a quantitative imaging biomarker of inflammation, this Section gives an overview of the physical principles of nuclear magnetic resonance (NMR), including precession, relaxation, signal detection and image generation. Additionally, quantitative MRI techniques for imaging inflamed bone – particularly diffusion-weighted imaging and chemical shift-encoded MRI – are considered in detail. These two methods underpin much of the work presented in this thesis.

2.3.2 Nuclear Magnetic Resonance

2.3.2.1 The origin of the magnetic moment

The basic building block of magnetic resonance imaging is the quantum mechanical property of spin, which describes an intrinsic form of angular momentum carried by elementary particles (it can be can be loosely interpreted as the particle spinning on its axis). Nuclei which have spin also possess a magnetic moment\(^3\), which can be

---

\(^3\) The magnetic moment of a magnet is a vector quantity that determines the torque (turning force) experienced by that magnet in a magnetic field.
considered by rough analogy with classical physics where the current associated with a rotating electrically charged body also generates a magnetic moment.

Although the uncertainty principle dictates that we cannot known the exact orientation of a particle’s angular momentum, we can determine a component of the angular momentum along a single direction. The component of the angular momentum \( I \) along an arbitrary direction (\( z \) is used by convention) is given by

\[
I_z = \hbar m
\]  

where \( m \) is the quantum number, which has a fixed number of possible values depending on the spin quantum number of the particle. In the case of hydrogen nuclei, \( m \) has two possible values and can be either \(-\frac{1}{2}\) or \(\frac{1}{2}\). Importantly, this means that \( I_z \) can also take two values.

In turn, the magnetic moment is proportional to the angular momentum and can itself take two values. The \( z \)-component of the magnetic moment is thus given by

\[
\mu_z = \gamma I_z
\]  

where the constant \( \gamma \) is known as the gyromagnetic ratio. When a magnetic field \( B_0 \) is applied, the energy of the spins depends on \( \mu_z \). Those protons which are aligned with the magnetic field occupy a lower energy position than those which oppose it.

The energy of an individual proton is given by the dot product of the magnetization vector and the static field, i.e.

\[
E = -\mu_z \cdot B_0
\]

Therefore, for a hydrogen nucleus, there are two available energy levels. Those protons aligned with the field are referred to as spin up (at the lower energy level), and those antiparallel to the field are referred to as spin down (at the higher energy level).

In the absence of a magnetic field, there is no difference in energy between the two energy levels (which are said to be degenerate) and there is no net magnetization in the tissue or object being imaged. However, in the presence of the \( B_0 \) field, the two
energy states are revealed, and there is a small preponderance of spins in the lower energy state (spin up), the size of which depends on a Boltzmann distribution

$$\frac{N_{up}}{N_{down}} = e^{\frac{\Delta E}{k_B T}}$$

where $\Delta E$ is the difference in energy between the two between spin states, $k_B$ is Boltzmann’s constant and $T$ is the temperature. The excess of spins in the spin up state is very small (typically in the order of 1 in 10,000), but this preponderance of spin up protons gives rise to the bulk magnetic moment for the tissue, which is the basis of MR imaging.

It should be noted that protons are able to transition between the spin up and spin down states. Quantum mechanical selection laws dictate that changes in $m$ can only take the value of $\pm 1$. From Planck’s law, the frequency association with a transition between energy states is

$$f = \frac{E}{\hbar} = \frac{\gamma B}{2\pi}$$

Where $f$ is expressed in Hertz. The frequency can also be expressed as an angular velocity

$$\omega = \gamma B$$

which is the well-known Larmor equation.

2.3.2.2 Behaviour of the bulk magnetic moment

In a population of spins, the bulk magnetic moment can be described in terms of classical mechanics. In a magnetic field, the bulk moment experiences a continuous turning force or torque which acts perpendicular to the spin axis. This torque causes the magnetic moment to precess around the strong magnetic field according to

$$\frac{d\mathbf{M}(t)}{dt} = \mathbf{M}(t) \times \gamma \mathbf{B}(t)$$

where $\mathbf{M}(t)$ is the bulk magnetic moment of the object being imaged, and $\mathbf{B}(t)$ is the applied field.
This relationship describes a movement known as *precession*, in which the axis of the magnetic moment itself rotates around the magnetic field (in a movement analogous to that of a spinning top), as shown in Figure 2-8.

![Figure 2-8 - Precession. The axis of the magnetic moment rotates around the magnetic field.](image)

This precessional motion can be mathematically described in several ways. One approach is to use Cartesian coordinates, in which $x$ and $y$ components of the magnetization each vary sinusoidally to produce the circular motion shown above.

An alternative approach is to describe the rotation in the $xy$ plane using *complex numbers*, where the magnetization vector $M$ in the $xy$ plane is expressed as the sum of a real and an imaginary number

$$M = x + iy.$$ \[8\]

as shown in the Figure below.

![Figure showing complex numbers](image)
Using complex notation, the magnetization vector can be expressed in terms of its angle (or argument) $\theta$ and its size $|M|$, i.e.

$$M = |M|e^{i\theta}$$ \[9\]

where the values $|M|$ and $\theta$ are known as polar coordinates. Using polar coordinates has the advantage that the $x$ and $y$ components do not need to be described separately, and the phase of the magnetization is represented using a single parameter, $\theta$.

The phase angle $\theta$ varies with time and depending on the frequency of precession, so the time-varying magnetization vector can be written

$$M(t) = M(0)e^{-i\omega t}.$$ \[10\]

where $\omega$ is the angular velocity, as per the Larmor equation.

This polar coordinate system is particularly well-suited to describing rotational movement and enables the motion of the magnetic moment to be described relatively simply and concisely. This system is particularly useful for describing the MR signal from tissues containing both fat and water and has been adopted in this thesis for the descriptions of chemical-shift encoded imaging (Section 2.3.6).

**2.3.2.3 Laboratory and rotating frames**

Description of the movement of the magnetic moment can be dramatically simplified by an appropriate choice of reference frame. The obvious choice is to define the reference frame relative to the scanner or magnet, such that the direction of the $B_0$ field is the $z$ direction and the $x$ and $y$ directions are perpendicular to the $z$ axis - this is known as the laboratory reference frame. However, the reference frame can also be defined such that $x$ and $y$ axes rotate around the $z$ axis - this is known as the rotating reference frame. If the frequency of rotation is chosen to match the frequency of precession, the rotation of magnetic moments will appear to have stopped and the magnetization can be visualized and analysed much more simply. Note that the axes in the laboratory frame are conventionally referred to as $x^\prime$, $y^\prime$ and $z^\prime$, whereas those in the rotating frame are referred to as $x$, $y$ and $z$ - this convention has been used in several of the Figures in the remainder of this Chapter. The process of converting
between the rotating and reference frames is described mathematically in Appendix A (Section 9.1.1).

The rotating frame concept is particularly useful for describing the action of a radiofrequency pulse on the object being imaged. When the frequency of the RF pulse matches the Larmor frequency (i.e. at resonance), an effective field is generated which is perpendicular to the magnetic moment in the rotating frame (which is designed such that the axes of the frame of reference rotate at the Larmor frequency) (171). This effective field causes the magnetization vector to undergo a ‘tipping’ movement (or, in other words, the magnetization precesses about the $B_1$ field), as shown in Figure 2-9.

![Diagram of magnetization before and after RF pulse](image)

**Figure 2-9 - Behaviour of magnetisation during an RF pulse.** At the start of the pulse ($t=0$), the magnetisation is aligned in the direction of the positive $z$-axis. During the pulse (at time $t$), the magnetization $\vec{M}(t)$ is gradually tipped in the $yz$ plane (rotating frame), with flip angle $\theta$. At the end of the inversion pulse ($t=T$), the magnetisation is aligned in the direction of the negative $z$-axis.

Under these conditions, the flip angle $\theta(t)$ during the RF pulse is simply proportional to the time integral of the $B_1$ field:

$$\theta(t) = \gamma \int_0^t B_1(t') dt'.$$  \[11\]
2.3.2.4 The Bloch equations

The Bloch equations (172) extend the framework for movement of the magnetic moment when relaxation is present. Specifically, they describe the additional contribution to the rate of change of the magnetic moment arising not from the applied magnetic field but from intrinsic processes in the tissue which contribute to dephasing.

Expressing all components individually, they can be written

\[
\begin{align*}
\frac{dM_x}{dt} &= \gamma (M \times B)_x - \frac{M_x}{T_2} \\
\frac{dM_y}{dt} &= \gamma (M \times B)_y - \frac{M_y}{T_2} \\
\frac{dM_z}{dt} &= \gamma (M \times B)_z - \frac{M_z - M_0}{T_1}
\end{align*}
\]

This motion can also be represented graphically, as shown in the following illustration. This shows the decay of transverse magnetization in the transverse plane, with regrowth of longitudinal magnetization in the z direction.

![Graphical representation of relaxation](image)

Figure 2-10 - A graphical representation of relaxation (simulated data in MATLAB) in the laboratory frame.

The physical processes underlying relaxation are also discussed briefly in Section 2.3.4.3, and have been described in detail elsewhere (173).
2.3.3 Image Formation

The previous section has explained the theoretical basis for excitation of spins using a radiofrequency pulse applied at the Larmor frequency. This section discusses how this process can be utilised to generate 3-dimensional images of an object or subject.

2.3.3.1 Spatial localisation

Slice selection

The first stage in spatial localization aims to excite only those spins within a fixed slice of tissue, so that any signal that is subsequently acquired can be ascribed to that slice. By successively sampling a number of these slices, a whole image volume can then be reconstructed.

The basic idea is to simultaneously apply a slice selection gradient, which imposes a linear variation of potential resonance frequencies along the object being imaged, and a specifically-tailored radiofrequency pulse, with frequency components matching the range of frequencies present in the desired imaging slice. The principle is demonstrated in Figure 2-11.

![Slice selection diagram](image)

Figure 2-11 - Slice selection. The RF pulse and slice selection gradient are applied simultaneously, enabling selective excitation of a specific tissue slice.
As shown in the Figure, the applied radiofrequency pulse is defined by a centre frequency and a frequency bandwidth (mathematically this corresponds to a sinc pulse in the time domain). In order to vary the thickness of the slice, the bandwidth of the pulse is typically held constant and the strength of the slice selection gradient is varied (a stronger gradient produces a thinner slice).

**Frequency encoding**

Once a slice has been selected from the imaging volume, further methods are required to form an image of that slice. The most common method is to apply a *frequency-encoding gradient*, such that each spatial location along the gradient direction in the object becomes associated with a unique resonance frequency, as shown in Figure 2-12. The acquired signal in the time domain will therefore consist of a range of frequencies, each corresponding to a known spatial location. A Fourier transform of the signal in the time domain is used to reveal the amplitudes which are present in the frequency domain.

![Frequency encoding](image)

Figure 2-12 - Frequency encoding.

The frequency encoding process can also be viewed in terms of the ‘collection’ of spatial frequencies present in the MR image. During the application of the frequency encode gradient, protons in the higher parts of the field will gain phase compared to those in the lower part of the field, thus creating a repeating pattern of phase shifts across the field of view (as in Figure 2-13). With time, the spatial frequency of the phase variation across the field of view will increase (the locations where the phase
exceeds 360 degrees – the phase \textit{wraps} - will become progressively more closely spaced). As a result, consecutive data points in the echo will correspond to progressively increasing spatial frequencies. When the spatial frequency at a specific time point matches a spatial frequency in the \textit{object}, the object will return a signal; when the spatial frequencies do not match then and there will be little or no signal returned.

A mathematical description of the signal acquired during frequency encoding is given in Appendix B (Section 9.1.2).

\textbf{Phase encoding}

To add another dimension of spatial encoding, the basic idea described in the previous section (frequency encoding) can be extended so that, in addition to the linear variation in frequency discussed previously, there is also a linear spatial variation in the \textit{phase} of the magnetization (which is perpendicular to the direction of frequency variation). This is usually achieved by applying an additional gradient lobe while the magnetization is in the transverse plane, but before the readout gradient. This introduces a linear phase variation which depends on position within the \textit{object}. Over subsequent imaging experiments, the size of this phase variation can be varied such that a separate spatial frequency is sampled for each experiment.
such that the whole image acquisition consists of a number of phase-encoding steps).

See also Appendix B (Section 9.1.2).

2.3.3.2 \textit{k-space}

The information acquired using frequency and phase-encoding methods is stored in a data format known as \textit{k-space}, as shown in Figure 2-14. \textit{k-space} consists of a matrix in which element corresponds to a pair combination of spatial frequencies and is the 2D or 3D Fourier transform of the MR image measured. A 2D/3D \textit{inverse} Fourier transform (which combines the contributions of all the different spatial frequencies) is used to derive the final image.

As the MR signal is complex (see Section 2.3.2.2 for an explanation of the complex nature of the MR signal), the result of this transform is a complex image with real and imaginary (or magnitude and phase) components. Most clinical imaging displays only magnitude images, but the phase of the signal does contain useful information which is utilized by some modalities. For example, the phase information is essential for accurate resolution of fat-water ambiguity in chemical shift-encoded MRI, which is discussed in detail in Section 2.3.6.

In a conventional Cartesian acquisition, each readout generates data to fill a single row of \textit{k-space}; the position of the row is defined by the phase-encoding gradient. See also Appendix B (Section 9.1.2) for further detail.
Figure 2-14 - k-space. Each phase encoding step (two individual phase encoding steps are labelled in red and blue) corresponds to a specific row in k-space. Each element corresponds to a specific pair of spatial frequencies—i.e. specific values for $k_{FE}$ and $k_{PE}$, as defined in Appendix A (Section 9.1). A 2D Fourier transform is used to reconstruct the image.
2.3.4 Creating Echoes

In MRI, the formation of an image relies on the generation of *echoes*, which are MRI signals generated following applied radiofrequency pulses. There are two main types of echo: gradient echoes (GE) and spin echoes (SE). In both sequences, a radiofrequency pulse is used for excitation, but the flip angle of this pulse is different for the two sequences. Spin echo sequences typically use 90° pulses whereas smaller angles are normally used for gradient echoes.

2.3.4.1 Gradient echoes

The principles underlying the gradient echo sequence are shown in Figure 2-15. Immediately after excitation (b), the bulk magnetization vector has been tipped into the transverse plane by the radiofrequency pulse. A negative gradient lobe is then applied after excitation (c), causing rapid dephasing of the transverse magnetization (much faster than the free induction decay of the MRI signal which would otherwise occur). Following this, a positive gradient is applied (d) which reverses the gradient and means that spins previously precessing at a low frequency now precess much faster. Therefore, spins in the image gradually rephase, and ultimately come back into phase and form a gradient echo (e).

![Figure 2-15 - Gradient echo sequence.](image-url)
The positive gradient reverses the dephasing caused by the preceding negative lobe but does not reverse dephasing caused by inhomogeneity in the magnetic field or due to spin-spin relaxation (discussed in the following Section), both of which will reduce the amplitude of the signal measured at (e). The acquired signal $S_{GE}$ therefore depends on the decay constant $T_2^*$, such that

$$S_{GE} = S_0 e^{-\frac{TE}{T_2^*}}$$  \[13\]

This decay constant $T_2^*$ can be split into an irreversible component, which arises from intrinsic relaxation processes in the tissue (see Section 2.3.4.3), and a reversible component, which arises from magnetic field inhomogeneity (this component is referred to as reversible because it can be eliminated or ‘reversed’ by using a spin echo rather than gradient echo sequence).

Accordingly, the rate of decay of the gradient echo signal ($R_2^* = 1/T_2^*$) is given by the sum of these two components, i.e.

$$R_2^* = R_2 + R'_2$$  \[14\]

where $R_2$ and $R'_2$ are the rate constants for irreversible and reversible dephasing respectively. It should be noted that the reversible component of the dephasing depends on the inhomogeneity of the applied field and also on local inhomogeneities which arise from tissue susceptibility (see Section 2.3.6.7); these local inhomogeneities are important in bone marrow imaging and contribute to an increase in $R_2^*$ (see Section 2.3.6.7).

### 2.3.4.2 Spin echoes

The spin echo sequence is shown in Figure 2-16. A key feature of this sequence is that an additional $180^\circ$ pulse is added before the readout gradients are applied.

A $90^\circ$ pulse initially flips the magnetization into the transverse plane (b), and the spins begin to dephase (c). The $180^\circ$ pulse then flips the spins through $180^\circ$ about the $y'$ axis so that their phase angles are reversed (d). Spins which were initially in a lower magnetic field now occupy a more advanced phase position compared to those which were initially in a higher magnetic field, and those which were in a higher magnetic field occupy a less advanced phase position. Over time, the spins gradually
rephase (e) until the spin echo is formed (f). Assuming that the spins have not moved during the imaging time, the reversible component of the dephasing (with the time constant $T'_2$) will disappear entirely at the spin echo. The decay is therefore dominated by the time constant for irreversible dephasing, $T_2$.

If diffusion effects are considered negligible, the equation can be simplified to

$$S_{SE} = S_0 e^{-\frac{TE}{T_2}}.$$  \[15\]

The physical processes which influence this time constant are discussed in the following subsection.

![Spin echo sequence](image)

**Figure 2-16 - Spin echo sequence.**

### 2.3.4.3 T2 and T1-weighting

**T2 and spin-spin relaxation**

The fundamental reason that spins dephase during an MR experiment is that there are subtle differences in the frequencies at which they precess, meaning that the bulk magnetization vector of the spin population reduces in magnitude. However, the reasons for dephasing vary: in the gradient echo experiment, relaxation is dominated
by the inhomogeneity in the magnetic field, whilst in the spin echo experiment relaxation is dominated by interactions between spins as they move within the tissues. This *spin-spin relaxation* (represented by the constant for irreversible dephasing, $T_2$) depends on the molecular motion (specifically the *tumbling rate*) of the protons in the imaged tissue and can therefore provide information about the structure of that tissue. If the molecules move freely (high tumbling rate) then the local magnetic field experienced by that molecule will be ‘averaged’ over a number of possible local fields, and the field will effectively be homogenous. However, if molecules are relatively fixed (low tumbling rate), the local magnetic field will be static and inhomogeneities within the tissue will therefore lead to relatively rapid dephasing. The relationship between the tumbling rate and the relaxation time $T_2$ is shown in Figure 2-17.

![Figure 2-17 - Relationship between tumbling rate and relaxation time.](image)

Importantly, unlike $T_1$ relaxation (see below) spin-spin relaxation does not involve any net loss of energy; the loss of net magnetization is due to dephasing alone.

The MR experiment can be *weighted* by the decay constant $T_2$ by choosing an echo time which is sufficiently long to allow tissues with different TEs to undergo
relaxation, but not so long that the signal from the tissues has almost reached zero. An intermediate choice of TE, designed to achieve optimal contrast between tissues of varying $T_2$, is shown in Figure 2-18. Alternatively, $T_2$ can be measured by repeating the MR experiment with different echo times, and then fitting an exponential function to the measured signal intensities.

![Figure 2-18 - Spin-spin relaxation and $T_2$ contrast. Contrast between tissues with varying $T_2$ can be optimized by an appropriate choice of TE, shown by the dashed grey line. The data shown were simulated using MATLAB.](image)

**$T_1$ and spin-lattice relaxation**

Spins can also lose energy through direct interaction with the surrounding tissues (the ‘lattice’) through a process known as **spin-lattice relaxation**. This energy can be absorbed via the tissues and dispersed through blood flow. Spin-lattice relaxation can be regarded as a process via which the protons approach thermal equilibrium with the lattice, which has a large heat capacity. Gradually, as the energy is transferred back to the tissue, the proportion of protons in spin up and spin down states reverts to its original value, and the magnetization along the $z$-axis recovers. This recovery due to spin-lattice relaxation is described the relaxation time $T_1$. Spin-lattice relaxation is much slower than spin-spin relaxation, taking several seconds compared to only a few hundred milliseconds for spin-spin relaxation.
The relative slowness of spin-lattice relaxation has important implications for MR sequence design. If the time between MR experiments (the repetition time, TR) is much longer than $T_1$, the next experiment will be unaffected because net magnetization along the z-axis ($M_z$) will have fully recovered before the experiment starts. However, if the TR is shorter, $M_z$ will not have fully recovered and there will be less available magnetization to flip into the transverse plane. The signal acquired from this subsequent experiment will therefore be reduced (weighted) depending on the $T_1$ value of the tissue being imaged. It can be shown that the z-magnetisation during an experiment is given by

$$M_z(t) = M_z(0)e^{-\frac{t}{T_1}} + M_0\left[1 - e^{-\frac{t}{T_1}}\right].$$

The repetition time (TR) can therefore be chosen such that contrast between tissues with varying $T_1$ is optimized, as shown in Figure 2-19.

In order for spins to relax from the spin down state (high energy) to the spin up state (low energy), those spins need to interact with other nuclei or molecules. This interaction is more efficient for spins which have a tumbling rate which is close to

![Figure 2-19 - Spin-lattice relaxation and $T_1$ contrast. Contrast between tissues with varying $T_1$ can be optimized by an appropriate choice of TR, as shown by the dashed grey line. The data shown were simulated using MATLAB.](image-url)
the Larmor frequency, which typically occurs for protons with intermediate binding (see Figure 2-17). Therefore, $T_1$ is shortest for protons with intermediate binding, but longer for both bound and free protons.

### 2.3.4.4 Fat suppression

In H\(^1\) MRI, the vast majority of the signal is derived from protons residing in fat and water molecules. In some tissues, fat contributes a large proportion of this signal and may therefore obscure pathology which specifically affects the water signal (this is particularly so in bone imaging where the normal marrow consists of approximately equal parts fat and water). Therefore, it is often desirable to separate or decompose the signals derived from fat and water. Fat protons have several properties which differ from those of water protons, which can be exploited using specific imaging techniques.

Firstly, protons in fat and water molecules resonate at slightly different frequencies, owing to subtle variation in their local chemical environments. These frequencies depend on the gyromagnetic ratio of the nucleus and on the magnetic field observed by the nucleus, which is in turn affected by the degree of electronic shielding of the nucleus. This frequency change is often described in terms of a ‘chemical shift’, $\sigma$, which modifies the Larmor frequency of a particular species of hydrogen atom according to

$$B_{\text{shifted}} = (1 - \sigma)B_0.$$  \[17\]

A positive value for the chemical shift indicates shielding, whilst a negative value indicates anti-shielding. The MR signal therefore consists of multiple spectral components which have frequencies separated by a few parts per million (ppm). In the case of triglycerides, the frequency of the main fat peak is shifted from that of water protons by approximately 3.5 parts per million (ppm).

Secondly, water and fat have different T1 values, owing to differences in molecular spacing and tumbling rates. MRI methods for exploiting these differences between fat and water are described in the following Sections.
**Fat suppression using chemical shift-selective excitation**

Chemical shift-selective (CHESS) imaging was first described by Haase et al. in 1985 (174). The CHESS method enables the selective excitation of specific species (e.g. fat or water); this allows the unwanted component (fat) to be removed whilst the desired component is unaffected.

Typically, a ‘saturation pulse’ is applied with a carrier frequency centred on the main fat peak, with a sinc-shaped or sinc-like RF amplitude envelope (175). The sinc-shaped pulse has a rectangular shaped in the frequency domain, producing approximately uniform delivery of RF energy across the main fat peak. Following the RF pulse, a crusher gradient is used to destroy the transverse magnetization. There is little time for the magnetization to recover before the remainder of the sequence itself begins, meaning that the fat-containing protons in the imaging volume contribute little signal to the final image. A schematic demonstration of this process, which is also known as fat saturation, is shown in Figure 2-20.

![Figure 2-20 - Fat saturation using CHESS.](image)

CHESS-based methods are effective if the $B_0$ and $B_1$ fields are homogenous but can perform poorly in areas of inhomogeneity. If the $B_0$ field is inhomogeneous, the RF pulse may be applied to a frequency band which is shifted relative to the fat peak, either missing the fat signal altogether or even suppressing the water signal. If the $B_1$ field is inhomogeneous, the fat signal may be only partially suppressed (even if
the RF pulse is applied at the fat frequency). Both of these problems mean that the residual fat signal (shown on the right-hand side of Figure 2-20) be substantial and have a significant impact on the final image. A high-quality shim is therefore essential for achieving uniform fat suppression.

It should be highlighted that the frequency shift between fat and water is proportional to the field strength of the magnet. In theory, $B_0$ inhomogeneity due to susceptibility effects should also double, meaning that the quality of fat suppression would not improve. However, in practice it is easier to suppress the fat signal without suppressing the water signal at higher field strengths, possibly because a wider spectral bandwidth of RF energy can be used, enabling shorter RF pulses in the time domain (175). These shorter pulses tend to perform better and may be easier to design (175). Conversely, at lower field strengths, longer RF pulses are needed to achieve a narrower RF bandwidth, necessitating some tradeoffs in the design process to achieve acceptable pulse durations.

A related idea is to use direct excitation of the water peak to achieve fat suppression, using ‘spatial-spectral’ pulses. These pulses can be used to excite both a specific slice of tissue and a specific spectral band (e.g. water) simultaneously. This method has been described in detail elsewhere (171,175).

**Fat suppression using inversion recovery**

The inversion recovery (IR) method for fat suppression was first proposed by Bydder et al. (176,177). The basic idea behind the IR method is to add an inversion pulse at the start of a spin echo sequence which is timed such that the fat signal reaches the zero point at the same time as the remainder of the spin echo sequence begins (Figure 2-21). When used for fat suppression, this method is generally referred to as *short inversion time inversion recovery* (STIR).

The behaviour of the magnetization during the inversion pulse can be described using the Bloch equations, as described previously. Although the aim of the inversion pulse is to rotate the magnetization vector by $180^\circ$, system imperfections mean that the actual rotation achieved is variable. Furthermore, if the spins are off-resonance, some transverse magnetization inevitably remains at the end of the inversion pulse. Inaccuracies in slice-selection gradients, chemical shift, magnetic field inhomogeneity and variations in susceptibility may all contribute to off-
resonance and therefore somewhat reduce the quality of the fat-suppression which can be achieved. Inversion pulses are commonly followed by spoiler gradients that dephase any residual transverse magnetization and therefore reduce the severity of artifacts in the final image.

Inversion recovery is a clinically practical method of fat suppression and is widely used. STIR images are a mainstay of clinical musculoskeletal imaging and are widely used to image inflammation (for example in the sacroiliac joint). T2-weighting can be introduced into STIR images by modifying the echo time of the spin echo component of the sequence, meaning that the images highlight areas of bone oedema very clearly. A major advantage of the STIR sequence is that it is relatively reliable in areas of $B_0$ inhomogeneity, which are otherwise difficult to image.

However, STIR sequences do have several disadvantages. The inversion recovery pulse introduces substantial T1-based contrast into the image, meaning that the observed contrast is a complex combination of T1- and T2-weighting. STIR

Figure 2-21 - The STIR sequence for fat suppression.
imaging should not be used with contrast agents (such as gadolinium) due to the risk of suppressing enhancing tissue (which will undergo T1 shortening in the presence of gadolinium and may therefore be closer to the null point at the time the 90° pulse is delivered). The STIR sequence is also relatively inefficient – the SNR of the water signal is reduced by around 50% compared to a conventional spin echo sequence (175).

CHESS- and STIR-based methods for fat suppression can be combined, with potential improvements in fat suppression. For example, the spectral presaturation with inversion recovery (SPIR) method uses a fat-selective RF pulse and spoiler gradient which is similar to CHESS, except that the RF pulse has a flip angle of 100-180°. This enables the use of an additional inversion delay mechanism, designed to ensure that any residual fat magnetization which was not nulled by the spoiler gradient is removed through the delay. The use of two spoiling mechanisms can potentially improve performance compared to a spoiler gradient alone.

The spectrally attenuated inversion recovery (SPAIR) method is similar to the SPIR method, except that the inversion pulse is different - an adiabatic pulse (which is relatively insensitive to B1 inhomogeneity) is used to selectively invert fat spins (171). Further detail on the radiofrequency pulses used in fat suppression is given in Ref (171).

2.3.5 Diffusion-weighted Imaging

Diffusion-weighted imaging (DWI) has been described and studied extensively in the field of neuroimaging, and detailed theoretical descriptions are given elsewhere (171,173,178,179). The aim of this Section is to give a brief overview of the principles of DWI with a view to specific applications in the field of bone imaging, and particularly to imaging inflammation.

2.3.5.1 Theory

The term molecular diffusion refers to the random displacement of molecules in a fluid due to agitation by thermal energy. This random motion can be statistically represented in terms of a displacement distribution, which describes the proportion of molecules that undergo displacement in a specific direction and to a specific
distance (180). Typically, the displacement distribution for free water molecules is Gaussian (Figure 2-22) (181).

The spread of this Gaussian distribution is described by a single parameter: variance ($\sigma^2$). In turn, the variance depends on the diffusion coefficient of the substance in question, $D$, which depends on viscosity or the ease with which molecules are displaced.

The value of $D$ can be calculated from the Stokes-Einstein equation, which describes the diffusion of spherical particles through a liquid (180):

$$D = \frac{k_b T}{6\pi \eta r}$$

where $k_b$ is the Boltzmann constant, $T$ is absolute temperature, $\eta$ is the dynamic viscosity and $r$ is the radius of the spherical particle.

In biological tissues, the tendency for water molecules to diffuse is limited not only by viscosity but also physical barriers – specifically cellular and intracellular membranes, intracellular fibres and large macromolecules (178,182). Intracellular proteins exert an important influence on intracellular diffusivity because they are charged, which causes protein-water adsorption (178). Importantly, the diffusion environment is very different in the intracellular and extracellular spaces: the intracellular space is generally a more ‘restricted’ diffusion environment than the
extracellular space, because of a greater density of macromolecules, fibres and membranes. Pathological processes which alter the balance between the intracellular and extracellular spaces therefore influence average diffusivity in the affected tissue (181). As such, measurement of diffusivity offers a window into the characteristics of a particular tissue at a micrometre scale (well beyond that of conventional MR imaging) (178, 183). Furthermore, diffusivity represents a property which is theoretically intrinsic to a particular tissue – if it can be measured, alterations in this property can be used to objectively evaluate disease.

### 2.3.5.2 Diffusion-weighted imaging sequences

In 1964, Stejskal and Tanner modified early spin-echo techniques using additional gradients to provide sensitization to diffusion (184). The Stejskal-Tanner sequence uses a T2-weighted spin echo sequence with two ‘motion-probing’ gradients (MPGs) on either side of the 180° refocusing pulse (Figure 2-23). Stationary water molecules within a voxel acquire a phase shift when the first MPG is applied, but acquire an equal and opposite phase shift at the time of the second MPG, such that their ‘net’ phase shift is zero. However, if a water molecule moves between the two MPGs due to diffusion, the second MPG does not completely compensate for the first and signal is attenuated.

![Figure 2-23 - The Stejskal Tanner sequence](image)

The degree of diffusion weighting imparted by the Stejskal-Tanner sequence depends on the duration of the MPG (δ), the interval between leading edges of
MPGs ($\Delta$), and on the strength of the MPG ($G$). An overall measure of the diffusion weighting, the *b*-value, can be calculated using the equation

$$b = \gamma^2 x G^2 x \delta^2 (\Delta - \frac{\delta}{3})$$ \[18\]

where $\gamma$ is the gyromagnetic ratio.

To determine the effect of the MPGs on the acquired signal intensity, we can consider the phase accumulation imparted by the MPGs. The phase accumulation $\phi$ imparted by the frequency shift $\Delta \omega$ at a specific spatial location $\vec{r}$ is given by

$$\phi = \int_0^t \Delta \omega \, dt' = \gamma \int_0^t \vec{G}(t'). \vec{r}(t') \, dt'$$ \[19\]

In the case of diffusion, the spins in a voxel are expected to move randomly over time, and it can be shown that the resultant signal intensity of a voxel is exponentially related to the variance $< \phi^2 >$ of a Gaussian distribution of phase values, which is equal to the product of the b-value and the diffusion coefficient $D$:

$$S = S_0 \cdot e^{-<\phi^2>} = S_0 \cdot e^{-bD}$$ \[20\]

where $S_0$ is the signal intensity on the T2-weighted (or $b=0$ s/mm$^2$) image, $D$ is the diffusion coefficient, and $b$ indicates the b-value (171).

DWI is usually performed with motion-probing gradients in three orthogonal dimensions, but can be performed with smaller or larger numbers of MPGs depending on the user’s purpose. If there is diffusional isotropy (equal limitation to diffusion in all directions) then it may be adequate to use a small number of MPG directions.

### 2.3.5.3 The apparent diffusion coefficient

Since the Stejskal-Tanner sequence is based on a T2-weighted spin echo sequence, the diffusion-weighted image is also T2-weighted (producing an effect known as *T2 shine-through*). To differentiate between T2 shine-through and genuine restricted diffusion, and to derive a quantitative measure of tissue pathology, images can be acquired at multiple b-values and the value for the diffusion coefficient $D$ is calculated. This can be achieved analytically or using a maximum likelihood (fitting)
approach depending on the number of data points which are acquired. When imaging biological tissues, the diffusion coefficient is often referred to as the apparent diffusion coefficient (ADC), since it reflects a number of tissue processes, including the presence of cellular membranes, rather than being a pure measure of the diffusivity of a specific substance. Assuming that fat suppression is used, the ADC value can be viewed as a snapshot of the behavior of water in a particular tissue.

2.3.6 Chemical Shift-encoded MRI

Chemical shift-encoded MRI (CSE-MRI) relies on the fact that fat and water protons resonate at slightly different frequencies, as described in Section 0. In a CSE-MRI experiment, the interference of the fat and water signals leads to oscillation in the signal over time, and repeated sampling of the signal at varying echo times enables separation of the fat and water signals (185). This separation enables CSE-MRI to be used as a method of fat suppression; indeed, it is a widely-used and practical technique. Additionally, the fat fraction (FF) - defined as the signal arising from fat protons divided by the sum of the signals from fat and water protons - can be calculated from the fat and water images. Like ADC, the fat fraction can be seen as a quantitative image-based indicator of biological and pathological processes – an imaging biomarker (186).

In this Section, the historical development of CSE-MRI methods is reviewed. Over time, CSE-MRI has evolved from a simple, two-point method (185) into a more sophisticated technique enabling correction for potential confounding factors such as T1 and T2* bias. With the newer ‘confounder-corrected’ techniques, the measured fat fraction can be referred to as a proton density fat fraction (PDFF), which theoretically indicates the relative concentration of fat and water protons in the tissue of interest.

2.3.6.1 Dixon’s method

The original chemical shift imaging method was proposed in the 1980s by WT Dixon (185). Dixon used a simple spin echo sequence in which the position of the 180° refocusing pulse was shifted (while the readout gradient was fixed) such that two images could be acquired: one at the spin echo (where water and fat were ‘in phase’, i.e. $t_1 = 0$), and a second with a TE shift:
\[ t_2 = \frac{1}{2f_F} \]  \[ ^{[21]} \]

where \( f_F \) is the chemical shift value for the main fat peak relative to water in Hertz. Separated water and fat images can then be ‘computed’ by simple addition and subtraction respectively:

\[
\rho_W = \frac{S_{IP} + S_{OP}}{2} \]
\[
\rho_F = \frac{S_{IP} - S_{OP}}{2} \]  \[ ^{[22]} \]

Dixon’s method was the first to enable the generation of separate water and fat images based on the different resonant frequencies of the two species, and represented a major technical advance at the time. However, the problem with

---

**Figure 2-24** - Fat-water ambiguity. In mixed voxels (a, containing both fat and water) interference at opposed phase echo times leads to loss of signal. In pure water (b) and pure fat (c) voxels, the magnitude of the signal is unchanged at opposed phase echo times, leading to ambiguity between pure water and pure fat, as in Dixon’s original method. However, phase data can be used to distinguish between these situations (i.e. resolve the fat-water ambiguity).
Dixon’s original approach is that imperfections in the applied field ($B_0$ inhomogeneity) and variations in magnetic susceptibility cause phase errors in the acquired in phase and opposed phase signals, leading to substantial leakage of signal between the fat and water images. Dixon eliminated the effect of $B_0$ inhomogeneity by taking the magnitude of both the in phase and opposed phase images before addition and subtraction (185). However, this approach is only accurate for water-dominant voxels (i.e. voxels where $\rho_w > \rho_f$) since the water image will always contain the largest of the two signal components, and the fat image the smallest (see Figure 2-24). Therefore, Dixon’s images contained an inherent ambiguity between high-fat and high-water tissues, which subsequent studies have attempted to address.

2.3.6.2 Resolving the ambiguity: extended two-point methods

To overcome the fat-water ambiguity problem, several authors extended Dixon’s method to make use of phase data (rather than magnitude alone), enabling the correct identification of high-fat and high-water voxels (187,188). These methods involve the acquisition of complex-valued images (i.e. images with both magnitude and phase data) at in phase and out of phase echo times (see Section 2.3.2.2 for an explanation of the complex nature of MR images). The phase information can be used to distinguish between high-fat and high-water voxels, as shown in Figure 2-24. However, these extended two-point methods are somewhat susceptible to the effects of $B_0$ inhomogeneity, since it can be difficult to differentiate between phase shifts due to $B_0$ effects and phase shifts due to chemical shift. One approach to reducing the effect of $B_0$ inhomogeneity is to combine these algorithms with phase-unwrapping or region-growing methods (fat-water separation is usually straightforward if the phase can be successfully unwrapped) (187,188).

Unfortunately, these phase-unwrapping algorithms are somewhat complex and tend to fail in areas of signal cancellation, particularly where the fat and water signals have similar magnitude.

2.3.6.3 Correcting for $B_0$ off-resonance: analytical three-point methods

To eliminate the need for complex region-growing methods and to improve estimation of $B_0$ inhomogeneity, several authors developed three-point CSE-MRI methods enabling direct estimation of field inhomogeneity to remove its effects from the signal.
With $B_0$ inhomogeneity included, the signal $S(t_n)$ in a given voxel can be written:

$$S(t_n) = (\rho_W + \rho_F e^{i2\pi f_B t_n}) e^{i2\pi f_B t_n}$$

where $f_B$ (in Hertz) is the local frequency shift due to static field inhomogeneity. $f_F$, the chemical shift of fat relative to water, is assumed to be known a priori whereas $f_B$ is unknown and is spatially varying. A schematic showing the additional contribution of $f_B$ to the fat and water signal vectors is shown in Error! Reference source not found..

Estimation of $f_B$ is made simpler by increasing the number of echo times acquired (189). A common approach is to use three echoes, as first described by Kim et al (190). Lodes et al. and Glover and Schneider both used schemes with echo times $t_1 = -T$, $t_2 = 0$ and $t_3 = T$ where $T = \frac{1}{2f_F}$, defined relative to the spin echo, which gives two opposed-phase images and one in-phase image (190–192):

$$S(t_1) = (\rho_W - \rho_F) e^{-i2\pi f_B T}$$  \[24\]

$$S(t_2) = (\rho_W + \rho_F)$$  \[25\]

$$S(t_3) = (\rho_W - \rho_F) e^{i2\pi f_B T}.$$  \[26\]
The off-resonance term $e^{i2\pi f_B t}$ can be obtained by dividing $S(t_3)$ with $S(t_1)$, which allows for calculation of $\rho_W$ and $\rho_F$. However, this calculation gives two possible solutions. Choosing the incorrect solution results in correct separation of fat and water but incorrect identification – in other words the water and fat will be swapped. A unique solution can be obtained by assuming that the field inhomogeneity is small (i.e. $|f_B| < \frac{f_F}{2}$), however, the assumption of small field inhomogeneity can easily be violated in practice because $|f_B|$ can be significantly larger than $f_F$. Glover and Schneider therefore proposed a method incorporating phase-unwrapping and a ‘phase-switch’ (192). The phase unwrapping algorithm uses the spatial derivatives of the measured phase to determine points at which the phase is wrapped and subsequently correct the phase image; in addition, trend analysis is used to deal with wraparounds beyond $2\pi$ (typically in regions of large localized field inhomogeneity). The calculated phase can then be used to determine the predominant component (i.e. fat or water) and therefore ‘switch’ the calculation to ensure correct fat-water assignment.

A disadvantage of the using symmetrically-acquired echoes – as in the method above (191,193) - is that fat-water decomposition is poor when the proportions of water and fat are approximately equal (194–196). Using symmetrically-acquired echoes is rather inflexible with regards to echo spacing. A more general version of the three-point method allowing greater flexibility (‘direct phase encoding’) was therefore proposed by Xiang and An (197), and authors such as Berglund et al. have subsequently adopted a similar approach (198). The signal from three images with uniformly-spaced TE shifts $t_0, t_0 + \Delta t$ and $t_0 + 2\Delta t$ can be expressed as

\[
S_1 = W + a_0 F
\]  \[27\]
\[
S_2 = (W + a_0 a F)b
\]  \[28\]
\[
S_3 = (W + a_0 a^2 F)b^2
\]  \[29\]

where $a_0$, $a$ and $b$ are complex phase factors (‘phasors’) due to chemical shift, defined

\[
a_0 = \exp (i2\pi \Delta f t_0)
\]  \[30\]
\[
a = \exp (i2\pi \Delta f \Delta t)
\]  \[31\]
\[ b = \exp(i2\pi\gamma B_0 \Delta t). \]  

Eqs. [27]-[29] can be used to generate a quadratic equation in \( b \)

\[ aS_1b^2 - (a + 1)S_2b + S_3 = 0, \]  

giving the solutions

\[ b_{1,2} = \frac{(a + 1)S_2}{2aS_1} \pm \sqrt{\left(\frac{(a + 1)S_2}{2aS_1}\right)^2 - \frac{S_3}{aS_1}}. \]

Again, only one of the two solutions for \( b \) can be correct. Several approaches have been used to resolve this ambiguity (197,198). For example, Berglund et al. showed that the correct solution could be obtained by determining the value of the ratio \( R = W/F \). For the correct value of \( b \), \( R \) will be a real non-negative number, whereas for the swapped solution \( R \) will have the same phase as \( (a_0 a)^2 \) (198). This method can be extended using a region-growing algorithm: starting with ‘seed’ voxels where the solution can be confidently determined, the value for \( b \) can be propagated to other voxels (198). Berglund’s method is fast and reported to be robust in areas of \( B_0 \) inhomogeneity, even in obese patients where \( B_0 \) inhomogeneity at the skin may be substantial.

### Maximum likelihood estimation

The analytical three-point methods described were only intended to separate two species (water and fat). In practice, the two-species assumption leads to inaccuracies in fat quantification, since human fat actually consists of a number of spectral fat components (the ‘main’ fat peak comes from hydrogen atoms in methylene residues, but methyl, allylic and olefinic residues also contribute substantially to the signal – see Figure 2-26, Table 3 and Figure 2-27) (199). This spectral complexity can cause a substantial deviation from the two-species assumption and therefore lead to inaccuracies in fat quantification if unaccounted for. Furthermore, three-point methods do not account for T2* decay, which can introduce a further source of bias (199,200). The effect of spectral complexity and T2* decay on signal evolution over time is shown in Figure 2-28.
Figure 2-26 - Single and multipeak fat models. In the multipeak model there are a number of fat peaks (shown schematically), each with a different chemical shift value, as shown in the Table below. This leads to additional complexity in the behaviour of the signal, as shown in Figure 2-27, below.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Chemical group</th>
<th>Chemical shift (ppm)</th>
<th>Relative amplitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methyl protons</td>
<td>=CH₃</td>
<td>0.90</td>
<td>0.087</td>
</tr>
<tr>
<td>Methylene protons</td>
<td>=-(CH₂)ₙ-</td>
<td>1.30</td>
<td>0.624</td>
</tr>
<tr>
<td>Methylene protons β to carbonyl</td>
<td>-CH₂-CH₂-COO</td>
<td>1.59</td>
<td>0.070</td>
</tr>
<tr>
<td>Methylene protons α to C = C</td>
<td>-CH₂-CH = CH-CH₂-CH₂</td>
<td>2.03</td>
<td>0.095</td>
</tr>
<tr>
<td>Methylene protons α to carbonyl</td>
<td>-CH₂-COO</td>
<td>2.25</td>
<td>0.067</td>
</tr>
<tr>
<td>Diallylic methylene protons</td>
<td>=CH-CH₂-CH =</td>
<td>2.77</td>
<td>0.016</td>
</tr>
<tr>
<td>Methine protons</td>
<td>=CH = CH-</td>
<td>5.31</td>
<td>0.042</td>
</tr>
</tbody>
</table>

Table 3 - Values for chemical shift of different fat components as measured by Ren et al. (201). The relative shown amplitudes shown are derived from Ren et al.’s data from marrow adipose tissue, and are almost identical to those from subcutaneous fat.
Figure 2-27 – NMR spectrum of subcutaneous fat (left) and tibial bone marrow (right), measured by Ren et al. (201). The bottom trace shows the acquired spectrum, and the upper trace shows the fitted spectrum.

Figure 2-28 – Simulation of signal evolution in the presence of spectral complexity of fat and T2* decay. Simulated signal/time plots are shown for (a) Single peak fat model with no T2* decay (b) Single peak fat model with T2* decay (c) Complex spectral fat model with no T2* decay and (d) Complex spectral fat model with T2* decay. The fat spectrum described by Ren et al. (201) was used for these simulations.
Inaccuracies in fat fraction estimates due to spectral complexity can be minimised using maximum likelihood estimation (MLE). Using this method, the signal model is pre-specified and the fitting algorithm finds the values of the parameters in the model which will minimize the error between the modelled data and the acquired data. MLE for fat-water decomposition was first proposed by An and Xiang, who used a non-linear least squares (NLLS) method with a signal model assuming multiple spectral fat components (202). Subsequently, a more general maximum-likelihood method allowing greater flexibility with echo times was proposed by Reeder et al, called echo asymmetry and least squares estimation (IDEAL) (203–205). A brief outline - based on Ref. (204) - is as follows.

The signal model in Eq. 3.23 can be extended to allow for multiple fat components

$$S(t_n) = (\rho_W + \sum_{m=1}^{M} p_m e^{i2\pi f_F,m t_n}) e^{i2\pi f_B t_n}$$

where \(p_m\) is the amplitude of each spectral fat component and \(f_F,m\) is the frequency offset of each fat component.

The unknown parameters in Eq. 3.35 are fitted using an iterative method involving two main steps which are each repeated in each iteration. In the first step, the unknown parameters are estimated with a linear least squares procedure, typically using 0 as a first guess for \(f_B\). In the second step, the error in \(f_B\) is calculated using a further least squares procedure, and a new estimate of \(f_B\) is obtained by adding the error to the initial guess. The first step is then repeated to re-calculate the parameters with the new estimate of \(f_B\). This procedure is repeated until the error in \(f_B\) (calculated in the second step) is small. Since estimating the field map at each voxel results in noisy estimates, IDEAL also incorporates a smoothing step, following which the smoothed field map can then be used to recalculate the water and fat amplitudes at each voxel. IDEAL can also be extended to incorporate T2* decay into the signal model, which substantially improves the accuracy of fat quantification in the presence of liver, and has been shown to be an accurate measure of fat content both in phantoms and in patients with hepatic steatosis (199,206) (204). The signal model becomes
IDEAL has several properties which make it attractive for use in clinical imaging. Assuming the correct initial estimate for $f_B$ is known, IDEAL results in the optimal fat-water decomposition at each voxel. Furthermore, it can be used for arbitrary echo times and arbitrary numbers of spectral components (207). Although IDEAL was initially described for spin echo acquisitions (where three echoes are acquired asymmetrically around the spin echo), it can be readily applied to gradient echo sequences. Using gradient echo imaging has the advantage that a larger number of echoes can be acquired relatively quickly - as a result, echo spacing is less crucial (204). Finally, the use of a larger number of echoes also allows for simultaneous $T_2^*$ estimation, which can be useful in diseases causing iron overload (204). IDEAL has formed the basis of a number of commercially-available algorithms for fat quantification designed for liver imaging, and has been widely used in medical imaging research (208–211).

However, a disadvantage of IDEAL is that it often fails in the presence of large field inhomogeneities, since the algorithm contains an implicit assumption that the field inhomogeneity is close to zero (the initial guess of $f_B = 0$ means that $f_B$ can only be directly determined if the true value is between $\pm f_B/2$). As a result, several ‘extensions’ have been proposed to guide the $B_0$ estimation process (once the correct field map is known, water fat separation is trivial, and reduces to a linear least-squares problem at every voxel). For example, Yu et al. developed a region-growing scheme which imposes field map smoothness on the image, starting with an automatically-selected seed voxel (212), and Tsao & Jiang proposed a ‘multiresolution’ method for field map estimation to help guide the decomposition (213). Region-growing and multiresolution methods can also be combined by using region-growing at the coarsest resolution, and then propagating the estimates to finer resolutions (214).

Although these methods represent improvements on the ‘voxel independent’ IDEAL method, they are arguably somewhat impractical and represent a significant limitation of the method. The fundamental problem with IDEAL is that the problem...
is ill-posed when voxels are considered individually, meaning that spatial regularization is required. A potential solution is to jointly estimate the complete field map for the entire image (215), which performs well in areas of high field-inhomogeneity and is robust in challenging situations (e.g. cardiac imaging) (215). Graph cut methods for whole image optimization are a focus of ongoing research (216).

2.3.6.5 Magnitude-based fitting

A potential simplification to the fitting process is to take the use only the acquired signal magnitude. This eliminates the need to estimate the field map, and also improves the signal to noise ratio (since the real and imaginary data are effectively repeat measurements). Taking the magnitude, Eq. [3.23] can be re-written

\[ |S(t_n)| = \sqrt{S_1^2e^{-2\nu_1t} + S_2^2e^{-2\nu_2t} + 2S_1S_2e^{-\nu_1t-\nu_2t}\cos(\omega t)} \]

where \( S_1 \) and \( S_2 \) are the water and fat components, and \( \nu_1 \) and \( \nu_2 \) are separate R2* terms for these components respectively (these are often combined into a single ‘system’ R2* term, \( \nu \)).

Unfortunately, using the magnitude creates an unavoidable ambiguity such that it is difficult to identify whether a voxel is fat-dominant or water-dominant – i.e. which of \( S_1 \) and \( S_2 \) corresponds to water, and which corresponds to fat – similar to the ambiguity in Dixon’s original method. One approach to resolving this ambiguity is to use an alternative source of contrast, such as T1 relaxation times (217). For example, Hussain et al. proposes a dual flip-angle (FA) approach to make use of the shorter T1 value for fat (as compared to water), consisting of two acquisitions with FA values of 70° and 20° (217). Depending on whether the estimated fat fraction increases or decreases with T1 weighting, each voxel can be designated as fat-dominant or water-dominant (217). This method provides a simple, quick method for resolving the ambiguity between water-dominant and fat-dominant voxels.

However, on theoretical grounds the algorithm is expected to perform poorly in voxels containing pure water or pure fat, since the measured fat fraction should not vary with flip angle in these voxels. Assignment of voxels as fat/water is therefore likely to be heavily influenced by noise (in the paper by Hussain et al., it is apparent
that the subcutaneous fat has a ‘speckled’ appearance, since this tissue has been labelled as mixture of fat-dominant and water-dominant FF values). The use of a dual flip angle acquisition also means that the fat-water decomposition is dependent on $B_1$-inhomogeneity.

A related approach is to use intrinsic differences in $T_2^*$ values for fat and water to resolve the ambiguity (fat has an effectively shorter $T_2^*$ than water due to its spectral complexity, see below), but in practice this method suffers from poor noise performance and generally does not result in adequate fat water decomposition (218).

### 2.3.6.6 The fat fraction as a biomarker

A number of CSE-MRI algorithms have been discussed in this Section. Importantly, the choice of method can have a significant impact on the accuracy of fat fraction measurements (accuracy can be defined as the ‘correctness’ of the measurement compared with a reference standard and minimize bias). For example, using maximum likelihood estimation with $T_2^*$ decay incorporated into the signal model (see above, Section 2.3.6.4) can help to remove or minimize $T_2^*$-related bias in FF measurements. Similarly, $T_1$-bias can be minimized by using a low flip angle (ideally $5^\circ$ or less) and long repetition times which minimizes the $T_1$-weighting of fat and water and therefore the impact on FF estimates. Other potential sources of bias - including the presence of multiple fat peaks, phase errors, noise, sampling error, $B_1$-inhomogeneity and J-coupling – cannot be removed entirely but can be minimized depending on the acquisition strategy and on the post-processing method (186,196).

If potential sources of bias have been minimized as far as possible, the fat fraction can be described as a *proton density fat fraction* (PDFF) – the ratio of the unconfounded fat signal to the sum of the unconfounded fat and water signals (219). Fat fraction measurements which are still biased by one or more of these factors may be referred to as signal fat fractions (sFF). In general, MLE-based FF estimates can be seen as PDFF measurements, while FF estimates using two- or three-point Dixon methods should be referred to as sFF measurements. Using PDFF rather than sFF potentially enables a more meaningful comparison of values between scanners, and across multiple sites. Recent studies evaluating PDFF measurements across multiple sites and manufacturers suggest that the reproducibility of
‘confounder-corrected’ PDFF measurements is excellent in the liver (220,221), although further work is required to examine reproducibility in other anatomical regions. Nonetheless, the reproducibility of PDFF measurements makes CSE-MRI an attractive quantitative technique for use in research and clinical practice.

2.3.6.7 CSE-MRI in the bone marrow

Susceptibility differences between trabeculae and bone marrow

Magnetic susceptibility indicates the degree of magnetization of a material in response to an applied magnetic field, and is given by

$$\mathbf{M} = \chi \mathbf{H}$$

where $\mathbf{M}$ is the magnetization of the material (i.e. the density of magnetic moments) and $\mathbf{H}$ is the magnetic field strength. The value of $\chi$ dictates whether a material is attracted to or repelled by a magnetic field, and in turn how that material will affect the local magnetic field surrounding it.

Performing CSE-MRI in cancellous bone is challenging because the bony trabeculae, which consist mainly of calcium, have different susceptibility values than the surrounding bone marrow, which consists mostly of fat and water. This difference in magnetic susceptibility causes inhomogeneities in the field (both within voxels and across the image). In a gradient echo experiment, this causes an increase in the rate constant $R_2^*$ (see Section 2.3.4 for definition of $R_2^*$) (222,223), which needs to be corrected for in order to avoid bias in FF measurements (224,225). Once $R_2^*$ decay is accounted for, there is good agreement between MRS-based and imaging-based PDFF values in the proximal femur and spine (224,225), and in ex vivo trabecular bone specimens filled with fat-water emulsions with known fat fractions (226).

An important consideration is whether to use a single ‘system’ $R_2^*$ for both water and fat components (199,200), or whether to use separate $R_2^*$ terms for water and fat (218). Although using separate $R_2^*$ decay terms for water and fat theoretically results in a more accurate fat-water separation (227), the use of a system $R_2^*$ term is arguably more practical and results in greater robustness to noise, and may be more suitable in short-$T_2^*$ environments (218). Phantom studies suggest that marrow fat quantification using a single $R_2^*$ term is accurate (226).
R2* and R2' as potential bone biomarkers
Apart from being used as a ‘correction factors’ in fat-water MRI experiments, R2* or R2' (where R2* = R2 + R2', see Section 2.3.4) measurements may be useful in their own right. It has been shown that there is an approximately linear relationship between R2*/R2' and BMD (223,228), suggesting a potential role in the assessment of osteoporosis. Accordingly, R2*/R2' have been shown to parallel apparent bone mineral density measured by dual-energy X-ray absorptiometry (DEXA) or quantitative computed tomography (QCT) (159,229–232). Kühn et al. showed that R2* measurements could discriminate normal and osteoporotic patients with reasonably high levels of sensitivity and specificity, despite concomitant variations in fat content in osteoporotic individuals (159). R2*/R2' values may provide specific information about the mechanical properties of bone (for example, bone mineral density and also yield stress and strength) (233). In theory, R2' measurements might give a purer measure of BMD since underlying variations in spin-spin relaxation are eliminated; however, implementing the necessary sequences is relatively impractical compared to multi-echo gradient echo imaging (234). Assuming that variations in R2 and relatively small compared to R2', clinical R2* measurements are likely to be dominated by R2'.

It should be noted that the effects of inhomogeneity are dependent on field strength; estimates suggest that R2* is approximately proportional to $B^{1.5}$ (234).

2.3.6.8 Phantoms for CSE-MRI
There are a number of studies demonstrating the use of test objects (‘phantoms’) in which the proportion of water and fat is systematically varied to validate MRI measurements of fat and water content. The complexity of these phantoms varies from simple mixtures of water and oil to complex objects incorporating real trabeculated bone in an attempt to create a more ‘physiological’ situation.

Acetone-water phantoms
A very simple phantom can be generated using a mixture of water and acetone (204). Acetone is soluble in water and exhibits a single spectral peak arising from the hydrogen atoms in its two methyl groups (shifted from the water peak by approximately 2.4 ppm); acetone-water mixtures therefore can be used as a simple two-component system. In fact, acetone-water systems are simpler than mixtures of water and fat, since biological fats consist of a number of different spectral
components. Superparamagnetic iron oxide (SPIO) contrast agent can also be added to acetone-water mixtures to shorten T2* and therefore simulate the effect of tissue iron (204).

**Fat-water phantoms**
Phantoms containing biologic fats and water can be generated using surfactants; the addition of thickening or gelling agents typically improves stability and lengthens the period over which the phantoms can be imaged without separation of the components. For example, Bernard et al. used carrageenan (a gelling agent which is widely used in the food industry) to stabilize mixtures of soya oil and water, which enabled generation of a full range of fat fraction values (i.e. 0 to 100%) (235). An alternative is to use gelatin or agar to create solid mixtures of fat and water (226) – this method creates very stable mixtures although it is not clear whether creating solid gels at high fat fractions (i.e. with mixtures consisting predominantly of oil) is feasible or reliable.

Test objects incorporating trabecular bone have been manufactured using trabecular bone cores from human cadavers or using specimens from knee/hip replacement surgery. A typical procedure is to remove marrow from the cores enzymatically and then fill the bone cores with mixtures of gelatin, water and peanut oil (226). Ideally, this approach requires the use of a vacuum chamber to eliminate air bubbles and a CT scanner to check that the cores are free from bubbles after manufacturing. An alternative method for investigating the effect of trabecular bone is to collect specimens of human vertebral bodies with varying bone density, and measure the bone density using CT (223).

**2.4 MRI Methods for Quantifying Inflammation**

**2.4.1 Conventional MRI**
From a theoretical perspective, proton MRI is well-suited to imaging bone inflammation, since the inflammatory exudate would be expected to water content and to increase the freedom of water diffusion (which impacts on both T2 and ADC measurements). However, there have been few studies relating histological changes to those observed on imaging. The presence of high signal on STIR sequences - commonly referred to as ‘bone marrow oedema’ - is used as a diagnostic criterion for active inflammation (119), but there is little evidence on this histological basis of this
lesion. Imaging criteria for inflammation in spondyloarthritis have therefore been developed empirically, and a diagnosis of ‘active’ inflammation on MRI does not necessarily have a direct pathophysiological correlate.

A suggested MRI protocol for use in spondyloarthritis has been detailed by ASAS, and consists of T2-weighted STIR and T1-weighted spin echo images of the sacroiliac joints (236). Specific features used to assess spondyloarthritis are discussed in the following Sections.

### 2.4.2 STIR Images Show Bone Marrow Oedema

The key imaging feature for diagnosis of spondyloarthritis on MRI is the presence of increased bone marrow signal on T2-weighted STIR images, with reduced signal on T1-weighted spin echo images (119). This signal change is also the basis for the SPARCC scoring system, which is commonly used in research to assess the severity of sacroiliitis (163). An example of a patient with extensive, bilateral bone marrow oedema in subchondral bone is shown in Figure 2-29.

![STIR T1W fsT1W + Gad](image)

Figure 2-29 - Extensive, bilateral, active inflammation (bone marrow oedema) of the sacroiliac joint in a patient with spondyloarthritis. There is high signal on the T2-weighted STIR image (a), low signal on the T1-weighted image (b), and high signal on a fat-saturated T1-weighted image after gadolinium injection (c).

Despite widespread acceptance of STIR hyperintensity as an indicator of active inflammation, the specific histological basis of this feature in spondyloarthritis is only partly understood, and there have been only a handful of studies investigating this issue.

To investigate the histological basis of bone oedema, Marzo-Ortega et al. reported biopsy findings with three patients with sacroiliitis as defined by the ASAS criteria
All three patients had normal plain radiographs but extensive bone marrow oedema on MRI (unilateral in two cases, bilateral in one case), and CT biopsy was performed to exclude a septic process (all were negative). All cases demonstrated an inflammatory infiltrate, whilst macrophages and osteoclastic polykaryons (i.e. multinuclear osteoclasts) were also seen. One patient also showed evidence of extensive marrow fibrosis.

In a larger study, Cui et al. performed CT-guided biopsies of the sacroiliac joints in 36 patients with spondyloarthritis, and conducted histological analysis of the biopsy specimens. Histopathological features indicative of spondyloarthritis included pannus formation/inflammatory cell infiltration, pathological changes in cartilage and/or subchondral bone, cartilage degeneration, and enthesitis. Imaging was performed using MRI, single-photon emission CT (SPECT), CT, and plain radiography. Using histopathology as the reference standard, SPECT, MRI, CT and plain radiographs had sensitivities of 92.8%, 96.4%, 73.3% and 64.2%, and corresponding specificities of 62.5%, 75%, 87.5%, and 87.5% for sacroiliitis. This study suggests that all of these modalities provide reasonably high levels of diagnostic accuracy for detecting sacroiliitis as detected using histology, although the lack of separation of specific features means that specific definition of bone marrow oedema (on imaging) in histological terms remains difficult.

On theoretical grounds, it is likely that STIR hyperintensities arise from a combination of factors, including oedema, cellular infiltration, changes in capillary permeability, changes in fat content and structural alterations. STIR contrast depends on both T1 and T2 values and the relative contribution of changes in T1 and T2 is not known. Given the nature of the inflammatory exudate, it seems likely that STIR hyperintensity is due to a combination of T2-lengthening (as a result of increased freedom of movement in the inflammatory exudate), and a proportional increase in water content in the marrow (due to exudation of water into the marrow, and possibly also a reduction in fat content). Areas of inflammation are typically hypointense on T1-weighted images, which would support the suggestion that the proportion of water is increased in inflamed tissue.

Active inflammation can also be demonstrated using a paramagnetic contrast medium (usually gadolinium), which shortens the T1 value and therefore results in hyperintensities on T1-weighted images in well-perfused areas. The information
provided by contrast-enhanced images is thought to largely overlap with that provided by STIR images (119). However, gadolinium may provide additional value in some situations, including in patients with early disease and in children (116,239) and for evaluating inflammation of the thoracolumbar spine (240).

Notably, Bollow et al. reported a significant association between disease activity measured using histology and the severity and type of inflammation on MRI, as assessed using a system based on fractional enhancement with gadolinium (38). However, the scoring system used in this study is not based on currently accepted clinical scores, and the relationships between histology and imaging findings were somewhat unclear.

2.4.3 T1-weighted Images Provide ‘Structural’ Information

There are several imaging abnormalities that are commonly visible in patients with spondyloarthritis even after treatment and are therefore described as a ‘chronic’ or ‘structural’ lesions. These lesions include fat metaplasia, erosions and joint fusion (ankylosis) and are usually assessed on T1-weighted images (see Figure 2-30).

*Fat metaplasia* is thought to represent an increase in the fat content of subchondral marrow (although this is unproven). Areas of fat metaplasia are defined by an increase in T1-weighted signal, and typically also show a reduction in signal on STIR images. It has been suggested that fat metaplasia results from fatty acid esterification due to inflammation (155), although again there is little evidence to support this idea.

*Erosions* are defined as bony defects in the joint margin (155). Erosions are typically regarded as chronic or structural lesions but can also be actively inflamed. They typically return low signal on T1-weighted images and high signal on STIR images when active and return progressively lower signal as they become chronic. If multiple erosions are present, these may become confluent and therefore cause ‘pseudodilatation’ of the sacroiliac joints.

*Joint ankylosis* (fusion) is subjective on MRI (which typically gives little information about bone mineral density), but described features include low signal on all conventional sequences and the formation of T1-hyperintense ‘bone bridges’ which cross the joint (155). When multiple bone bridges are present, the joint cavity may
become blurred (this also makes consistent identification of fusion subjective and rather inconsistent).

2.4.4 Quantitative Methods for Imaging Inflammation

To date, the main quantitative techniques investigated in spondyloarthritis have been diffusion-weighted imaging and dynamic contrast-enhanced MRI, although most studies have been small and have used rather inconsistent methodology.

Studies using diffusion-weighted imaging to quantify inflammation have been discussed in the following Chapter (Section 3.1).

DCE-MRI was initially described by Tofts et al., and involves the collection of a series of images after contrast injection, enabling the generation of an enhancement curve for each pixel (241). This method is, in theory, well-suited to imaging inflammation since increased tissue perfusion is central to the inflammatory process. In a study by Zhang et al., perfusion parameters including $K_{\text{trans}}$, $K_{ep}$, $V_e$, time to peak (TTP), maximum concentration and area under the curve (AUC) were

Figure 2-30 - Chronic inflammatory (‘structural’) lesions in spondyloarthritis. (a,b): Bilateral erosions of the sacroiliac joints (solid red arrow) are shown, surrounded by areas of fat metaplasia (blue arrow). (c,d): There is bilateral ankylosis of the sacroiliac joints (open red arrow) and extensive fat metaplasia (blue arrow).
significantly higher in patients with active sacroiliitis compared to those with inactive sacroiliitis (242). Using a receiver operating characteristic (ROC) analysis to determine the ability of these parameters to discriminate active and inactive disease showed that $K_{\text{trans}}$ had the highest AUC of the parameters investigated, at 0.836. Gaspersic et al. also showed that semi-quantitative DCE parameters such as enhancement factor ($f_{\text{enh}}$) and gradient ($g_{\text{enh}}$) were sensitive to reductions in inflammatory activity over time, although their analysis was restricted to single inflammatory lesions (243).

Unfortunately, the need for contrast injection is a significant downside of DCE-MRI, due to the associated cost, pain and anxiety (particularly in children) associated with the injection and the risks of gadolinium deposition (244,245). The reproducibility of quantitative DCE-MRI parameters such as $K_{\text{trans}}$ is also notoriously poor, potentially reducing the sensitivity of these parameters (246). As a result, this thesis has focused on methods that do not require contrast injection for the evaluation of inflammation.
3 The Apparent Diffusion Coefficient as a Biomarker of Inflammation

3.1 Overview

This Chapter deals with the use of DWI to measure inflammation in spondyloarthritis. In Section 3.2, previous studies using DWI in spondyloarthritis are reviewed. In Section 3.3, a new method for measuring ADC values in the sacroiliac joint, which was developed in our group, is discussed. Section 0 describes a retrospective study evaluating the apparent diffusion coefficient as a marker of response to biologic therapy. The studies presented in this Chapter were performed retrospectively and were used as the starting point for the subsequent work presented in this thesis.

3.2 Background

In spondyloarthritis, DWI has been proposed as a method for quantifying inflammatory activity, and for monitoring disease severity over time. Multiple studies have shown that ADC measures are elevated in areas of bone marrow oedema compared to normal marrow (243,247,248). This is thought to be due to the presence of the inflammatory exudate, which increases water content, increases the size of the extracellular space (where water molecules diffuse more freely) and may cause a proportional reduction in fat content at inflamed sites.

It should be highlighted that most studies using DWI in spondyloarthritis have focused on the sacroiliac joints (SIJs). The SIJs are a good ‘imaging model’ of skeletal inflammation in general since they are large and immobile, and therefore relatively easy to image. Additionally, the SIJs have particular clinical importance as sacroiliitis is a cardinal feature of spondyloarthritis, and examining these joints clinically is difficult as they are deep and cannot be palpated. To date there have been few or no studies using qMRI to image other joints in spondyloarthritis, but it seems likely that the inflammatory process – and the associated imaging features – would not vary dramatically between anatomical locations.

The first study using DWI in spondyloarthritis was performed by Gaspersic et al., who followed individual areas of inflammation over time (so-called ‘inflammatory
lesions’) and showed significant reductions in ADC in patients treated with infliximab (243). Subsequently, Bozgeyik et al. found that ADC values in patients with sacroiliitis (n=13) were significantly higher than those in controls with mechanical back pain (n=29) (247). More recently, Gezmis et al. performed a study of 62 patients with clinically definite spondyloarthritis and 40 controls, and showed that ADC measurements performed on both surfaces of the sacroiliac joint (for both joints) were significantly higher in patients than in controls (248).

A recent advance in the use of quantitative DWI is the development of methods which separately evaluate an early pseudo-diffusion coefficient (\( D_{\text{fast}} \)) and tissue diffusivity (\( D_{\text{slow}} \)) (249). In theory, these methods allow separate evaluation of perfusion abnormalities and bone marrow oedema respectively. Zhao et al. showed that \( D_{\text{slow}} \) measurements were significantly higher in patients with active sacroiliitis than in chronic sacroiliitis and controls, whereas \( D_{\text{fast}} \) measurements were not significantly different between groups (249). This result suggests that an increase in water diffusivity, rather than an increase in perfusion, may be the major contributor to the increased ADC in areas of bone marrow oedema. However, this research is at an early stage and there have been no attempts to rigorously validate any of these parameters as imaging biomarkers thus far.

In a qualitative sense, DWI has also been investigated as alternative to contrast-enhanced images for detecting oedema (with the aim of avoiding the need for contrast injection) (250). It is currently unclear whether DWI provides any additional benefit over conventional imaging for the purposes of qualitative assessment (251).

### 3.3 The Linear ROI method

#### 3.3.1 Rationale

If ADC measurements are to be clinically useful as a measure of inflammation severity, it is essential to provide an overall, ‘summary’ measurement for each patient which can be used to guide management decisions. However, previous studies (discussed above) have typically only followed individual areas of inflammation over time, or used circular ROIs placed in specific areas of subchondral bone, and none have attempted to derive an overall inflammatory measure or score (243,247,248). Furthermore, the diffusion protocols used previously have not been
standardized, meaning that ADC measurements may vary depending on the specific 
b-values used and on the quality of fat suppression. This lack of standardization may 
impact on the reproducibility of ADC measurements across scanners, and on the 
ultimate utility of ADC as an inflammatory biomarker.

Therefore, there is a need for a technique for measuring ADC which gives an overall 
inflammatory score for each patient, and which gives reliable measurements across 
scanners. The linear ROI technique (252) was developed for this purpose. This 
method can be used to produce overall mean ADC measurements for each patient 
and can also derive a normalized ADC value, which theoretically could be more 
reproducible than the uncorrected ADC (253–256).

3.3.2 Description of Method

The linear ROI method was developed in our group with the aim of providing 
objective, patient-specific measures of inflammatory severity (252). In this method, 
three linear ROIs are placed across the synovial portion of each sacroiliac joint (as 
shown in Figure 3-1), for four consecutive ADC slices, producing 24 ROIs in total. 
Each linear ROI generates a ‘profile’ of ADC values. The simplest way to analyse 
these profiles is to simply take an overall mean ADC value from all 24 of these ROIs 
(this can be referred to as the ‘raw’ or uncorrected ADC value). The ADC value can 
also be normalized by placing a further ‘reference’ ROI on normal sacral 
interforaminal bone to provide internal standardization; the normalized ADC 
(nADC) value is defined as mean uncorrected ADC divided by the reference ADC. 
Both ADC and nADC measurements can potentially be used as a measure of 
inflammatory activity (in theory, nADC measurements might be advantageous 
because of the internal standardization provided by the reference ROI).

As a proof-of-principle, the linear ROI technique was evaluated in 10 patients with 
spondyloarthritis and 10 controls with mechanical back pain (252). nADC values 
were found to be significantly higher in cases than in controls (p=0.0015), and there 
was also a strong correlation between STIR scores and nADC values (Figure

* The tool was developed by David Atkinson and Kani Vendhan; this author performed ADC 
measurements and drafted the final manuscript for the initial proof-of-principle study for this tool. 
The subsequent ‘biological validation’ study included in this Chapter was performed primarily by 
this author; full author lists for both publications are given in (Section 0)
This study indicated that using the linear ROI technique is feasible, and provides ADC/nADC measurements which reflect inflammation of the sacroiliac joints. However, as highlighted by Sullivan et al. (1), a number of further steps are required to validate ADC as a biomarker of inflammation. The following study aims to address one of these steps – demonstrating that the biomarker reflects biological change in patients undergoing treatment.

Figure 3.1 - Linear ROI technique. Placement of linear regions of interest (ROIs) (red) is shown on the apparent diffusion coefficient (ADC) map of a patient with sacroiliitis (a) and a control with mechanical back pain (b). Only one slice is shown for each patient (four slices were scored in total). (c) The reference ROI is placed on interforaminal sacral bone (blue line) and used to generate a reference profile. (d) ADC profiles from the two subjects (in red) are shown compared with the reference profile from interforaminal sacral bone (in blue). The ADC values are higher for the patient with sacroiliitis (left) than for the normal subject (right).

Figure 3.2 - Preliminary results showing nADC measurements against visual inflammation scores.
3.4 ADC as a Marker of Response to Therapy

3.4.1 Introduction

The initial study by Vendhan et al. proposed the use of ADC and nADC as patient-specific measures of sacroiliac joint inflammation, which might be used to monitor disease activity over time (252). This study showed that nADC is significantly correlated with inflammation as measured using conventional visual scoring (252). However, there is currently no evidence on whether these measurements are sensitive to biological change, which is essential to demonstrate if an imaging biomarker is to be used for treatment monitoring (1). Here, a retrospective study was performed to assess ADC and nADC as measures of response in young patients with spondyloarthritis undergoing TNFi therapy.

It was hypothesized that the change in ADC and nADC after treatment would be greater in responders to TNFi therapy than in non-responders to TNFi therapy.

3.4.2 Methods

3.4.2.1 Subjects

A local clinical adolescent rheumatology database was used to identify all those patients with sacroiliitis aged between 12 and 30 years who had been started on biologic therapy between January 2009 and June 2015. A picture archiving and communication systems (PACS) search was then performed to identify those individuals who had undergone MRI scans of the SIJs between both before and after starting biologic therapy, using a pre-defined imaging protocol as specified below (see Image Acquisition). Patients who started on biologic treatments during this period and who had MRI scans both before and after starting therapy were selected for the study. All subjects fulfilled the International League of Associations for Rheumatology (ILAR) criteria for ERA (69) and were treated with either etanercept or adalimumab. The decision to scan the patients and to treat with biologic therapy was made as part of standard clinical care in all cases. At our institution, ERA patients are typically scanned at presentation, 3 months after starting treatment (usually methotrexate) and again after a further 3-6 months if patients are started on biologic therapy to confirm improvement of inflammatory changes (67). A subset of patients are also scanned at regular intervals (typically yearly) following this for disease monitoring (67).
3.4.2.2 Image Acquisition

MRI of the SIJs was performed using a 1.5T Siemens Avanto scanner. Diffusion-weighted imaging was performed in a single shot using an echo planar readout, with b-values of 0, 100, 300, 600 and 1000 s/mm²; fat suppression was performed using SPIR. Other acquisition parameters included TR/TE 3500/87, FOV 316mm, matrix size 128x128, pixel size 2.5mm, slice thickness 8mm, averages 4, slices 17, EPI factor 120. ADC maps were generated on vendor software using a standard monoexponential fit. Conventional clinical MRI sequences, consisting of axial STIR (TR/TE 6070/83ms, TI 150ms, slice thickness 5mm), coronal turbo inversion recovery magnitude (TR/TE 610/11ms, TI 150ms, slice thickness 4mm), axial and coronal T1-weighted (TR/TE 475-610/11ms, slice thickness 3 and 5mm respectively) and axial and coronal post-contrast T1-weighted images with SPIR fat suppression (TR/TE 619-795/11ms, slice thickness 3-5mm) were also acquired, in keeping with ASAS recommendations (119).

3.4.2.3 Image Analysis

nADC measurements were performed using the linear ROI tool (252), as follows. For each of the central four axial slices on the ADC maps, three linear regions-of-interest (ROIs) measuring 14mm in length were manually drawn across the synovial portion of each SIJ, with each ROI centred on the joint space. Where the AP dimensions of the joint were too small to place three ROIs, only two ROIs were placed. A further ‘reference’ ROI was placed on normal sacral bone to provide internal standardization. The normalised ADC (nADC) value of each patient was defined as the ratio between the mean ADC of all joint line profiles and the mean reference ADC. ROIs were not placed on areas of the joint where the images were affected by fat ghosting artifact (these were typically visible as low signal ‘bands’ in the image with a similar morphology to subcutaneous fat, arising due to unsuppressed fat shifting across the image).

For each scan, both the ‘uncorrected’ ADC and nADC were recorded. The measurements were performed independently by two radiology registrars (KV and TB, with seven and three years of musculoskeletal MRI experience respectively); the mean of the two radiologists’ scores was used for the analysis.
The change in ADC after therapy ($\Delta$ADC) was defined as:

$$\Delta\text{ADC} = \text{ADC}_{\text{pre}} - \text{ADC}_{\text{post}}$$

and the change in nADC after therapy was:

$$\Delta\text{nADC} = \text{nADC}_{\text{pre}} - \text{nADC}_{\text{post}}$$

Note that positive $\Delta$ADC and $\Delta$nADC values represent a reduction in the post-treatment nADC (i.e. improving inflammation).

To establish a reference standard, the SPARCC technique for visual scoring of bone marrow oedema on STIR images (163) was modified for use on axial rather than coronal images, as previously described (252). Specifically, on each of the central six axial slices, the SIJ was divided into four quadrants. Increased STIR signal was given a score of 1 per quadrant and normal signal was scored 0. For each slice, an additional score of 1 per joint was given for deep or intense lesions (163). Each patient received a maximum score of 12 per slice, and a maximum total of 72. Scoring was performed independently by two radiologists (MHC and TB) with over twenty years and three years of musculoskeletal MRI experience respectively, who were blinded to clinical response classifications and to the ADC/nADC measurements. The mean score from the two sets of measurements was used for the analysis. The change in visual score on STIR images after therapy (referred to here as $\Delta$STIR) was defined as:

$$\Delta\text{STIR} = \text{STIR}_{\text{pre}} - \text{STIR}_{\text{post}}$$

### 3.4.2.4 Response Classification

Radiological response classification was based on changes in visual score after treatment. Specifically, based on previous studies defining a ‘minimally important change’ for visual scores of sacroiliac joint inflammation (257), patients were classified as **radiological responders** if the mean score from the two observers improved by 2.5 or more, and **radiological non-responders** otherwise. This threshold was derived using a receiver operating characteristic (ROC) analysis to determine the visual score change which would predict ‘minimally important’ changes in inflammation (as determined by a radiologist) with the highest degree of sensitivity and specificity (257).
Each patient was classified as either a clinical responder or clinical non-responder to TNFi therapy using a retrospective physician global assessment (PGA). Specifically, a specialist consultant adolescent rheumatologist (NA) who was blinded to visual, ADC and nADC scores reviewed the electronic medical record to determine clinical symptoms at the time each scan was acquired (i.e. both the pre-TNFi and post-TNFi scans) to define a composite global assessment of response to treatment. Patients who required emergency steroid treatment (defined as a course of systemic steroids to treat a flare – i.e. oral prednisolone, intramuscular or intravenous methylprednisolone) or a switch to an alternative TNFi at the time of the second scan were classified as non-responders. Patients who had only marginally improved [defined as an improvement in PGA of less than 30%, mirroring the PGA component of the ACR Pedi 30 criteria (258, 259)] were also classified as non-responders. Patients whose letters contained insufficient detail to enable a response assessment were excluded.

3.4.2.5 Statistical Analysis

The Mann-Whitney-Wilcoxon test was used for between-group comparisons. The correlation between change in $\Delta$ADC/$\Delta$nADC and $\Delta$STIR was evaluated using Spearman’s rho. Receiver operating characteristic analyses were performed to assess sensitivity and specificity for determining response using both $\Delta$ADC and $\Delta$nADC, using both clinical and radiological response classifications. Inter-observer variability was assessed separately for pre-treatment and post-treatment nADC and ADC measurements using intraclass correlation coefficient (ICC) and Bland-Altman 95% limits of agreement.

3.4.3 Results

3.4.3.1 Demographics

Twenty-two patients were recruited, with a mean age at biologic initiation of 17y4m. Eighteen subjects were male (mean age 17y3m), and four were female (mean age 17y6m). The mean interval from pre-treatment scan to the initiation of biologic therapy was 4m (range 1m to 8m). The mean interval from start of biologic to post-treatment scan was 1y1m (range 5m to 2y6m).
3.4.3.2 Disease Response: Radiological Classification

Of the 22 patients in the cohort, 18/22 patients (82\%) demonstrated a STIR score improvement of $\geq 2.5$ and were classified as radiological responders; 4/22 (18\%) were classified radiological non-responders. Uncorrected ADC and nADC values before and after treatment, classified by radiological response, are shown in Table 4 and Figure 3-3.

Baseline ADC values were significantly higher pre-treatment in radiological responders compared to non-responders ($p=0.03$). After treatment, there was a decrease in ADC values in responders and an increase in non-responders, such that post-treatment ADC values were higher in non-responders than in responders ($p=0.16$). Furthermore, the change in ADC values ($\Delta$ADC) was significantly greater in responders than in non-responders ($p<0.01$).

Baseline nADC values were also higher in responders than in non-responders, although this difference was non-significant ($p=0.31$). Following treatment, nADC values were marginally lower in responders than in non-responders, again non-significant ($p=0.22$). There was a reduction in nADC values after treatment in both responders (median $\Delta$nADC = 0.27) and non-responders (median $\Delta$nADC = 0.10). The change in nADC values (i.e. $\Delta$nADC) was greater in responders than non-responders; this difference was borderline-significant ($p=0.055$).

3.4.3.3 Disease Response: Clinical Classification

Of the 22 patients in the cohort, 18/22 patients (82\%) were classified as clinical responders, and 4/22 (18\%) were classified as clinical non-responders. Of the clinical non-responders, two patients had also been classified as radiological non-responders. ADC and nADC values before and after treatment, classified by clinical response, are shown in Table 5 and Figure 3-4.
Chapter 3 - The Apparent Diffusion Coefficient as a Biomarker of Inflammation

Table 4 - Median (SD) ADC, ΔADC, nADC and ΔnADC values in radiological responders and non-responders. The two groups were compared using the Mann-Whitney-Wilcoxon test: p-values are shown in the right-hand column. Positive ΔnADC and ΔADC values reflect an improvement in inflammation.

<table>
<thead>
<tr>
<th></th>
<th>Radiological responders</th>
<th>Radiological non-responders</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADC pre-treatment (mm²/s x 10⁻⁶)</td>
<td>912 (219)</td>
<td>654 (204)</td>
<td>0.03</td>
</tr>
<tr>
<td>ADC post-treatment (mm²/s x 10⁻⁶)</td>
<td>696 (208)</td>
<td>962 (359)</td>
<td>0.16</td>
</tr>
<tr>
<td>ΔADC (mm²/s x 10⁻⁶)</td>
<td>+240 (342)</td>
<td>-308 (204)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>nADC pre-treatment</td>
<td>1.67 (0.68)</td>
<td>1.29 (0.77)</td>
<td>0.31</td>
</tr>
<tr>
<td>nADC post-treatment</td>
<td>1.42 (0.30)</td>
<td>1.75 (0.52)</td>
<td>0.22</td>
</tr>
<tr>
<td>ΔnADC</td>
<td>+0.21 (0.75)</td>
<td>-0.25 (0.38)</td>
<td>0.055</td>
</tr>
</tbody>
</table>

Figure 3-3 - Response plots for nADC and ADC by radiological response. Pre- and post-treatment nADC and ADC values are shown for all 22 patients, classified according to radiological response. Patients whose nADC/ADC reduced between after treatment are shown in green, whilst patients whose nADC/ADC increased are shown in red. ADC values have units mm²/s x 10⁻⁶.
Table 5 - Median (SD) ADC, ΔADC, nADC and ΔnADC values in clinical responders and non-responders. The two groups were compared using the Mann-Whitney-Wilcoxon test; p-values are shown in the right-hand column. Positive ΔnADC and ΔADC values reflect an improvement in inflammation.

<table>
<thead>
<tr>
<th></th>
<th>Clinical responders</th>
<th>Clinical non-responders</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADC pre-treatment (mm²/s x 10⁻⁶)</td>
<td>897 (254)</td>
<td>849 (163)</td>
<td>0.90</td>
</tr>
<tr>
<td>ADC post-treatment (mm²/s x 10⁻⁶)</td>
<td>675 (215)</td>
<td>905 (334)</td>
<td>0.097</td>
</tr>
<tr>
<td>ΔADC (mm²/s x 10⁻⁶)</td>
<td>+139 (372)</td>
<td>-102 (425)</td>
<td>0.33</td>
</tr>
<tr>
<td>nADC pre-treatment</td>
<td>1.61 (0.73)</td>
<td>1.65 (0.57)</td>
<td>0.77</td>
</tr>
<tr>
<td>nADC post-treatment</td>
<td>1.39 (0.29)</td>
<td>1.79 (0.29)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>ΔnADC</td>
<td>+0.21 (0.78)</td>
<td>-0.12 (0.34)</td>
<td>0.089</td>
</tr>
</tbody>
</table>

Figure 3-4 - Response plots for nADC and ADC by clinical response. Pre- and post-treatment nADC and ADC values are shown for all 22 patients, classified according to clinical response. Patients whose nADC/ADC reduced between after treatment are shown in green, whilst patients whose nADC/ADC increased are shown in red. ADC values have units mm²/s x 10⁻⁶.
There was no significant difference in baseline ADC values between clinical responders and non-responders \( (p=0.90) \). Post-treatment ADC values were higher in non-responders, although this difference was again non-significant \( (p=0.097) \). There was no significant difference in the change in ADC values \( (\text{i.e. } \Delta\text{ADC}) \) between responders and non-responders \( (p=0.33) \).

There was no significant difference in baseline nADC values between responders and non-responders \( (p=0.77) \). Following treatment, nADC values were significantly lower in responders than in non-responders \( (p<0.01) \). Accordingly, there was a reduction in nADC values in clinical responders \( (\text{median } \Delta\text{nADC} = 0.21) \) and an increase in nADC values in non-responders \( (\text{median } \Delta\text{nADC} = -0.12) \); the difference between these groups was borderline-significant \( (p=0.089) \).

### 3.4.3.4 Receiver operating characteristic analysis

ROC analyses for nADC and ADC, for clinical and radiological response, are shown in Figure 3-5.

**Using radiological criteria for response classification:**

*nADC:* Any decrease in nADC after treatment was 89% sensitive and 75% specific for distinguishing radiological responders \( (\text{i.e. those with a reduction in SPARCC score } \geq 2.5) \) from radiological non-responders. The area under the ROC curve \( (\text{ROC AUC}) \) was 0.82, with a sensitivity of 89% and specificity of 75% at the optimal operating point.

*ADC:* Any decrease in ADC after treatment was 67% sensitive and 100% specific for distinguishing radiological responders from radiological non-responders. ROC AUC was 0.97, with a sensitivity of 100% and specificity of 75% at the optimal operating point.

**Using clinical criteria for response classification:**

*nADC:* Any decrease in nADC after treatment was 83% sensitive and 50% specific for distinguishing clinical responders from clinical non-responders. ROC AUC was 0.78, with a sensitivity of 95% and a specificity of 50% at the optimal operating point.
ADC: Any decrease in ADC after treatment was 50% sensitive and 50% specific for distinguishing clinical responders from clinical non-responders. ROC AUC was 0.67, with a sensitivity of 95% and a specificity of 50% at the optimal operating point.

3.4.3.5 \( \Delta \text{ADC}, \Delta n\text{ADC} \) as Continuous Response Measures

Figure 3–6 shows the relationship between change in ADC/nADC (\( \Delta \text{ADC} \) and \( \Delta n\text{ADC} \) respectively) and change in SPARCC STIR score (\( \Delta \text{STIR} \)) after TNFi treatment. There was a significant, positive correlation between \( \Delta \text{ADC} \) and \( \Delta \text{STIR} \) (R=0.60, p=0.031) and between \( \Delta n\text{ADC} \) and \( \Delta \text{STIR} \) (R=0.55, p<0.01).
3.4.3.6 Inter-observer Variability

For ADC, the pre-treatment ICC for interobserver variability was 0.98 (Bland-Altman 95% limits of agreement $0.09 \pm 105 \times 10^{-6}\text{mm}^2/\text{s}$) and the post-treatment ICC was 0.96 (Bland-Altman 95% limits of agreement $14.0 \pm 142 \times 10^{-6}\text{mm}^2/\text{s}$). For nADC, the pre-treatment ICC was 0.93 (Bland-Altman 95% limits of agreement $-0.06 \pm 0.52$) and the post-treatment ICC was 0.81 (Bland-Altman 95% limits of agreement $0.09 \pm 0.43$).

3.4.4 Discussion

Here, we found that changes in ADC and nADC measurements were significantly higher in responders to TNFi therapy than in non-responders, as defined using standard radiological criteria. Changes in ADC and nADC were also higher in clinical responders than in clinical non-responders, but this difference was non-significant ($p=0.33$ and $0.089$ for ADC and nADC respectively). Changes in ADC/nADC were positively correlated with changes in visual inflammation score, and could be used to classify patients as responders and non-responders. These data suggest the ADC and nADC reflect changes in inflammation in patients undergoing TNFi therapy, and could potentially be used to monitor response to treatment. ADC and nADC measurements acquired using the quantification tool described here are potentially more objective than visual scoring, and do not rely on the opinion or expertise of the observer to identify inflammation (although using this method the observer does need to place the ROIs). Using a quantitative approach could therefore help to improve the reproducibility of inflammation measurement, and help to rationalise therapeutic decision making in the clinic. Quantitative measurements of inflammation could be particularly useful in centres which lack specialist radiological expertise or where the caseload of inflammatory diseases is relatively small, and could also act as objective endpoints in clinical trials.

Importantly, ADC and nADC were more accurate predictors of radiological response than of clinical response, and we did not find a significant difference in $\Delta$ADC or $\Delta$nADC values between clinical responders and non-responders. This may be because clinical assessment is only an indirect measure of inflammation and is also influenced by various biomechanical, psychological and social factors – by contrast, STIR scoring and ADC/nADC measure inflammation (against which
TNFi treatment is directed) in a relatively direct fashion. Accordingly, previous studies have found clinical assessment to be an insensitive tool for diagnosing sacroiliitis (105) and that JIA patients in clinical remission frequently have evidence of ongoing inflammation on MRI scans (260). Occult inflammation could be prognostically important because of the potential for structural damage and fusion (115,261) which contribute to disability (262).

Although the results of this study do suggest that ADC/nADC measurements reflect response to treatment, it is important to note that there were a number of cases where changes in ADC/nADC were discordant with the clinical and/or radiological response. 2/22 patients and 6/22 patients for nADC and ADC underwent increases in these parameters despite being classified as responders using conventional radiological scoring. Although some discordance is expected between nADC/ADC measurements and clinical response (due to the complex nature of the clinical response assessment), discordance between nADC/ADC and radiological response may be due to limitations of the linear ROI technique. The three ROIs placed on each joint (on each slice) are only likely to include a small proportion of the total subchondral bone, potentially creating a sampling error. This sampling error may be particularly severe for the reference ADC measurements, where only one ROI is placed on normal subchondral bone. Accordingly, in a separate study (included in Appendix A, Section 10.1), we found that nADC values had significantly poorer inter-observer reproducibility than ADC values, which was largely attributable to variability in the reference ADC measurement (263). Similarly, the fact that the ROI crosses the joint may limit the agreement between ADC/nADC measurements and visual scores, since the SPARCC system for inflammation assessment only includes oedema in subchondral bone (inflammation in the joint space is not included) (163). For future work, it would be preferable to develop a method measuring all of the subchondral bone, but excluding the joint space.

There are also some general limitations of DWI as a quantitative technique which are important to highlight. Firstly, ADC measurements in the bone marrow suffer from relatively poor reproducibility: a recent review by Dietrich et al. summarized 24 studies which have measured ADC values in normal bone marrow and showed a wide variation in the values obtained (most of studies included showed ADC values between 0.2 – 0.6 x 10^-3 mm^2/s, whilst a smaller number found values up to 1.6 x 10^-3 mm^2/s) (264). This variability may be partly due to differences between pulse
sequences used at different centres and in the specific b-values which are used. The quality of fat suppression can also vary between scanners and can potentially influence ADC measurements substantially, since fat ADC values are much lower than water ADC values (264–266). This effect might be particularly important in areas of bone inflammation where there may be heterogeneous areas of oedema and fat metaplasia, which would be expected to cause variation in the measured ADC irrespective of the true water diffusivity (115). Areas with higher fat content would be expected to have a lower ADC value, due to increased contamination of the water signal. This issue has not been specifically investigated to date, although anecdotally it is apparent that ADC values are lower in areas of fat metaplasia. This point is demonstrated in Figure 3-7. A related problem is the presence of fat ghosting artifacts in diffusion-weighted images (267). Fat ghosts are thought to arise due to unsuppressed fat ‘shifting’ across the image in the phase-encoding direction, exacerbated by the low bandwidth of the EPI readout (see Figure 3-8) (267). These artifacts can be subtle and are therefore not always recognized by the observer, but can nonetheless result in underestimation of ADC measurements from affected sites. One potential solution is to switch the phase-encoding direction from anteroposterior to right-left (Figure 3-8), which ‘moves’ the artifacts away from the joint.

Improvements in the reproducibility of ADC measurements might be achieved by standardization of imaging protocols and development of calibration tools which can be used to adjust for measurement drift between sites. Some groups have also investigated methods for improving the quality of fat suppression by suppressing the olefinic peak (which is normally unaffected by the SPAIR or SPIR prepulse) using Dixon-based methods. For example, Hernando et al. described a method relying on an echo time-shifted spin echo acquisition (6 TE shifts), which enables magnitude-based separation of the water and residual olefinic fat in each voxel (268). More recently, Burakiewicz et al. developed a similar method for imaging muscle, again magnitude-based separation of water and olefinic fat but enabling the use of shorter echo times (through the use of partial Fourier acquisitions in the diffusion-weighted images) (269). It should be noted that both of these methods account for the spatial displacement of the olefinic fat signal in the models used to separate the water and fat signals.
Figure 3-7 - Effect of fat metaplasia on the apparent diffusion coefficient. Areas of fat metaplasia (solid arrow) and normal marrow (dashed arrow) are shown on T1-weighted (a) and STIR (b) images, and ADC maps (c). ADC values are lower in areas of fat metaplasia than in normal marrow.

The retrospective nature of the study has also imposed some limitations on this work. The sample size was small due to the relative rarity of the disease and the lack of availability of pre- and post-treatment scans, and the small number of non-responders (n=4) is likely to have reduced the power of the study. We also had no control over the time interval between patients starting TNFi therapy and the second scan. Spondyloarthritis is a chronic disease and SIJ changes are expected to occur over a long timescale, but it would be preferable to scan the patients at a fixed interval after starting treatment (preferably three or six months post-TNFi).

Finally, clinical response classification was performed retrospectively, which limits the accuracy of response measurement. It would be desirable to classify response prospectively using established tools in future work. Nonetheless, the fact that both clinical and radiological response assessments yielded similar results is a strength of the analysis and suggests that the data are likely to be meaningful.

3.5 Conclusion

Diffusion-weighted imaging is a conceptually attractive method for imaging bone inflammation, since the freedom of water diffusion should be directly altered in
inflammatory exudates. The linear ROI tool offers an objective method for quantifying inflammation, potentially offering a means to avoid the subjectivity associated with visual scoring. Here, we found that ADC and nADC reflect changes in inflammation in patient undergoing TNFi therapy, suggesting that these measurements could be used to monitor the effects of treatment. However, both the linear ROI method and DWI in general suffer from some technical limitations, and further methodological research is needed to overcome these issues. Prospective and ideally larger studies with definitive response classifications are required to validate ADC and nADC measurements more conclusively.

3.6 Acknowledgements

Versions of work included in this Chapter have been published in two articles, with the following references. Kani Vendhan and this author contributed equally to the first of the two studies (KV and DA designed the linear ROI tool and developed the analysis plan; TB and KV performed the measurements, and TB performed the statistical analysis and drafted the manuscript for submission). The second of the two studies was performed primarily by this author, as per the author details in the references below.


†Denotes equal contribution
Figure 3-8 - Fat ghosting artifacts in DWI. Two representative slices (A and B) are shown from a single patient. For each slice, b=0 images (top row) and ADC maps (bottom row) are included. (a,b): Fat ghosts cover the right ilium, and would therefore alter quantitative measurements. (c,d): Switching the phase encoding direction to RL alters the location of the artifacts and would therefore minimize their impact. However, as shown in (g,h), this switch can cause artifacts in other slices.
4 Rationale for Using CSE-MRI and Method Optimisation

4.1 Overview

In the previous Chapter, diffusion-weighted imaging was discussed as a method for imaging inflammation. Although this method is conceptually attractive and has produced some promising initial results, it has several technical limitations. In this Chapter, an alternative quantitative imaging method – chemical shift-encoded MRI (CSE-MRI) - is considered. This method has the potential to avoid several of the problems inherent to DWI, and has some particular properties which make it well-suited to imaging bone inflammation. The rationale for using CSE-MRI to image bone inflammation is considered in detail in Section 4.2. In Section 4.3, in order to select and optimize and specific CSE-MRI method for use in subsequent studies, a number of candidate algorithms were evaluated using example datasets from two volunteers. This evaluation led to the identification of two algorithms which have been used in subsequent work presented in this thesis, and also for further studies currently being conducted by our group.

4.2 Rationale for using CSE-MRI to image bone inflammation

The existing literature indicates that CSE-MRI has several properties which make it attractive quantitative method, and suggest that it could be superior to diffusion MRI as a quantitative modality in inflamed bone. Perhaps most importantly, CSE-MRI can accurately measure the proportion of water and fat in tissue (199), whereas ADC measurements are influenced by fat content in a complex, non-linear fashion which depends on the quality of fat suppression, the presence of fat-ghosting artifacts and probably also on the specific b-values which are chosen (see Section 3.4.4). In patients with spondyloarthritis, where inflammatory changes are often mixed (potentially causing both increases and decreases in fat content at different sites within the subchondral bone), this is a significant advantage. Additionally, the reproducibility of FF measurements across sites and across vendors is excellent (220). In the liver, ICC values as high as 0.999 have been reported in a study evaluating reproducibility across multiple sites, vendors, field strengths and imaging protocols (220). Having a truly reproducible biomarker effectively increases the diagnostic accuracy (1), and may increase the likelihood of uptake into clinical practice.
Another advantage is that the spatial resolution of CSE-MRI is typically superior to that of DWI - on a 3T Ingenia scanner, we have found that CSE-MR images can be acquired with a voxel size of 2 x 2 x 2mm in approximately 5 minutes, whereas DW images would typically have a slice thickness of 5-8mm and a larger voxel size with a similar acquisition time. CSE-MRI may also be less susceptible to artifacts than DWI (although it can suffer from fat-water swaps, improvements in algorithm technology have significantly improved the quality of fat-water decomposition) (198,205,270).

The fact that normal marrow has a FF of around 50% means that there is scope for both increases and decreases in FF, which could be used to measure both acute and chronic components of the inflammatory process (oedema and fat metaplasia). This might mean that active and chronic inflammation could be evaluated simultaneously using a single acquisition, thereby reducing scan time.

Finally, gradient echo-based CSE-MRI acquisitions are relatively fast, meaning that this method could easily be extended to whole-spine or whole-body imaging. This could enable relatively fast, quantitative assessment of the inflammatory burden in patients with spondyloarthritis or other diseases causing skeletal inflammation [for example, chronic relapsing multifocal osteomyelitis (CRMO)].

4.3 Optimisation of CSE-MRI for Imaging Bone Marrow

4.3.1 Overview

Having discussed the rationale for using CSE-MRI to image inflammation and the background theory in the previous section, the work presented in this Section aims to identify and optimize a suitable CSE-MRI algorithm which can be used for the studies subsequently presented in this thesis and beyond. A series of ‘candidate’ CSE-MRI methods were applied to two subjects’ image datasets, and were evaluated in terms of the following properties:

1. **Quality of fat-water decomposition.** This is determined by the correct assignment of high-fat fraction values to fat-dominant tissues, and of low fat fraction values to water-dominant tissues. If fat-water swaps are present, these should be restricted to the edge of the field of view.
2. **Accuracy of fat fraction measurements** compared to reference standard measurements. This can be assessed using phantom measurements as a reference standard or using spectroscopy; here, we have adopted the former approach.

3. **Ability to correct for $R_2^*$ decay.** This is an intrinsic property of the algorithm in use; $R_2^*$ correction is likely to be important in the bone marrow where the presence of trabeculae will increase the rate of dephasing.

An additional consideration - less easy to quantify but equally important - is ease of use. This includes whether fat-water decomposition can be performed in line (on the scanner) and, if offline processing is necessary, the time needed to process the images.

### 4.3.2 Evaluation of Candidate Methods

#### 4.3.2.1 IDEAL

The IDEAL method (which was described in Section 2.3.6.4) has been viewed as a ‘gold-standard’ method for fat quantification in the liver, and has been shown to be accurate in phantom experiments and using spectroscopy as a reference standard. The IDEAL implementation described by Yu et al. (204) can be implemented with gradient echo data, and corrects for $R_2^*$ decay. This algorithm was therefore investigated first.

**Methods**

Raw complex data were acquired from two healthy volunteers on a Philips 3T Ingenia scanner using a six-echo spoiled gradient echo sequence with monopolar readout (TE1 1.18ms, ΔTE 1.6ms, flip angle 3°, TR 25ms, pixel spacing 1.56 x 1.56, matrix size 320 x 320, slice thickness 2mm, bandwidth 1894 Hz/Px). These data were processed offline using an in-house implementation of the IDEAL algorithm described by Yu et al. (204), which accounts for $R_2^*$ decay (the code written to implement the algorithm is included in Appendix A, Section 9.2.1). Briefly, the following steps were used:

1. An initial guess of 0 was used for the complex field map for all pixels
2. The corresponding estimates for water and fat were obtained using a least squares inversion
3. Error terms are obtained using a further least squares inversion
4. The complex field map is obtained and the process is repeated (15 iterations were used in this work).

FF maps were generated from fat and water images using the signal magnitudes, and were evaluated by this author using a qualitative scale to assess the severity of fat-water swaps (No swaps – correct assignment of fat and water in all pixels, Minor swaps – incorrect assignment in less than 10% of pixels, Moderate swaps – incorrect assignment in 10-25% pixels, Severe swaps – incorrect assignment in more than 25% of pixels).

**Results and Discussion**

![Example fat fraction maps from the two subjects’ datasets, generated using IDEAL.](image)

FF maps from the two subjects’ datasets are shown in Figure 4-1. There are severe fat-water swaps in both subjects. Most of the tissues are misassigned – the muscles have been incorrectly labelled as high fat, whilst the subcutaneous fat has also been incorrectly assigned, with a fat fraction around 20%.

Although the precise reasons for failure of the IDEAL algorithm in these datasets is uncertain, the problem is likely to relate to inaccuracies in the estimation of the complex field map (218). This work adopted the approach in the initial IDEAL studies of using an initial guess of zero for the complex field map (203–205); however, it is known that an incorrect initial guess can cause convergence to local minima during the estimation process, and subsequent swaps. Implementations of IDEAL using an initial guess of zero (as used here) can be therefore be regarded
‘voxel-independent’ in the sense that the fitting for each voxel (particularly the fieldmap estimate) is not influenced by adjacent voxels (204,205). Following the initial description of IDEAL, a number of authors have since worked on ‘voxel-dependent’ methods incorporating region-growing algorithms which may improve estimation of the complex field map (224,225,271). These methods may offer improvements in fat-water separation (and a reduction in fat-water swaps), although these techniques can still be prone to inconsistent performance, are computationally-intensive, and may not be available in some centres.

**Conclusion**
As implemented here, the voxel-independent IDEAL algorithm resulted in severe fat-water swap artifacts. This implementation of IDEAL is therefore not adequate for clinical use and will not be used for the studies presented in this thesis.

### 4.3.2.2 Magnitude-based Reconstruction

Magnitude-based CSE-MRI (which was described in Section 2.3.6.5) is a relatively simple method which eliminates the need for complex field map estimation, and could therefore be used as an alternative to IDEAL. This method enables estimation of (and correction for) $R_2^*$ decay in a relatively straightforward fashion, since $R_2^*$ terms can simply be added to the signal model. Although magnitude-based fitting does not resolve the ambiguity between water-dominant and fat-dominant pixels, this ambiguity can potentially be resolved using a separate algorithm or by repeating the acquisition at multiple flip angles (217). Therefore, the magnitude-based method was first evaluated *without* attempting to resolve this ambiguity; a potential method for resolving this ambiguity is addressed in the following Section.

**Methods**

*Assessment of image quality in volunteers*

To assess image quality, raw complex images from the two volunteers described in the previous Section were processed offline using a nonlinear least squares fitting algorithm, with the signal model

$$|S(t)| = \sqrt{S_1^2 e^{-2v_1 t} + S_2^2 e^{-2v_2 t} + S_1 S_2 e^{-v_1 t - v_2 t} \cos(\omega t)}$$
as previously described by Bydder et al (218). The code written to implement the magnitude fitting algorithm is included in Appendix A, Section 9.2.2. Fitting was performed using the MATLAB ‘lsqcurvefit’ function, which uses a trust-region-reflective method for cost function minimization. The fat and water signals $S_1$ and $S_2$ (not necessarily in that order) were given equal start points of 100, which was chosen empirically. The $R_2^*$ decay term was a given start point of 0.1. FF maps were generated using the fitted signal magnitudes, and were assessed qualitatively to assess the severity of fat-water swaps using the qualitative scale described above.

**Interleaved multiecho acquisitions in an acetone water phantom**

In order to characterize the evolution of fat and water signals over time, and to assess the accuracy of magnitude-based fat fraction measurements, an acetone-water phantom was constructed and imaged using CSE-MRI (in an acetone-water phantom the acetone is used as a simple surrogate for fat, with a single spectral peak; see also Section 2.3.6.8). Acetone-water mixtures consisting of varying proportions of acetone [SpeedyPlastics] and distilled water, such that acetone fractions varied between 0% and 100%, in 10% increments. Acetone fractions in the phantom were calculated by volume, and adjusted acetone fractions were calculated to allow for differences in proton density between acetone and water (as shown in Table 6).

![Figure 4-2 – Acetone-water phantom. The percentages shown refer to the acetone fraction by volume; adjusted acetone fractions were used for the analysis.](image)

<table>
<thead>
<tr>
<th></th>
<th>Acetone</th>
<th>Water</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Density (kg/m³)</strong></td>
<td>791</td>
<td>1000</td>
</tr>
<tr>
<td><strong>Molecular weight (kg/mol)</strong></td>
<td>0.0581</td>
<td>0.018</td>
</tr>
<tr>
<td><strong>Protons</strong></td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td><strong>Moles/m³</strong></td>
<td>13614</td>
<td>55556</td>
</tr>
<tr>
<td><strong>Protons per m³/protons per m³ of water</strong></td>
<td>0.735</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 6 - Adjustment for proton density of acetone.
Data from the phantom were acquired using a series of interleaved acquisitions in which the first TE varied but the echo spacing was kept constant, allowing data collection at very finely spaced echo times. Specifically, thirty separate echo times were acquired from 1ms to 6.75ms using five interleaved six-echo acquisitions, with bipolar readout. On each subsequent acquisition, the first echo time was increased (TE1/ΔTE were 1ms/1ms, 1.2ms/1ms, 1.4ms/1ms, 1.6ms/1ms and 1.75ms/1ms for the five acquisitions). The repetition time was 25ms, and a SENSE factor of 2 was used to accelerate the acquisition. Images were processed using the magnitude-based fitting algorithm described above (details in Appendix A, Section 9.2.2). The chemical shift parameter ω was allowed to vary freely between 1.3 and 4 kHz, and assigned a start point of 2 kHz. All other parameters had a lower bound of 0. Again, S1 and S2 were given equal start points of 100, which was chosen empirically, and the R2* term was a given start point of 0.1.

To assess the effect of varying flip angle on the accuracy of fat fraction measurements, the acquisition was performed with multiple flip angles (2°, 5°, and 15°).

**Results and Discussion**

*Assessment of image quality in volunteers*

FF maps from Subjects 1 and 2 are shown in Figure 4–3. As expected, fat-water swaps were observed in all fat-dominant voxels (for example, the subcutaneous fat). These swaps appeared to be entirely attributable to the unresolved fat-water ambiguity using the magnitude-based method, and no further artifacts were observed.

*Interleaved multiecho acquisitions in an acetone water phantom*

The acetone-water phantom was constructed successfully. Examples of the signal oscillations observed in the mixtures are shown in Figure 4–4. In all of the mixtures (i.e. all those tubes other than pure water and pure acetone), the goodness-of-fit was excellent (all R-squared values were greater than 0.99). In pure acetone and pure

---

As discussed in the results, we unexpectedly found a gradual reduction in the size of the chemical shift between acetone and water with increasing acetone concentration. This may be due to hydrogen bonding between acetone and water in solution (272,273). By allowing ω to vary, potential bias introduced by this effect could be minimized.
water tubes, the fit was less accurate (R² values were <0.30) although the measured acetone fraction agreed closely with the reference standard in these tubes (Figure 2). The poorer R² values in these ‘pure’ tubes may be due to signal oscillations arising from the bipolar readout gradients (there is a visible difference in signal intensity between odd and even echoes in the pure acetone and pure water tubes).
Unexpectedly, we observed a gradual reduction in the magnitude of the observed chemical shift with increasing acetone concentration, as shown in Figure 4-5 (it is also apparent in Figure 4-4 that the in-phase and out-of-phase echo times vary depending on acetone fraction). This may be due to concentration-dependent variations in the nature of hydrogen bonding between acetone and water. Studies of acetone-water mixtures suggest that addition of water to acetone introduces hydrogen bonding with the carbonyl group of the acetone molecule \((272,273)\). Since the oxygen atoms of both water and acetone can accept either zero, one or two hydrogen bonds from water there are a number of possible acetone-water complexes, which may change structure depending on concentration \((273)\). Previous magnetic resonance studies have also demonstrated that individual T1 values for water and acetone in water-acetone mixtures vary depending on the composition of the mixture, again suggesting there are significant water-acetone interactions in solution \((274)\).

Failure to consider these effects could lead to inaccuracies in quantification when using acetone phantoms for quality control (particularly if the value for \(\omega\) is fixed in the fitting algorithm).

![Figure 4-5 – \(\omega\) plotted against acetone fraction over the range of acetone fractions in the phantom. In all the tubes containing a mixture of water and acetone (i.e. acetone fraction 0.1 to 0.9), there was a gradual reduction in \(\omega\) with increasing acetone fraction.](image)
Measured acetone fractions have been plotted against true acetone fractions in each vial in Figure 4-6. Above 50%, there is unresolved fat water ambiguity, meaning that acetone fraction measurements are swapped (i.e. the measured acetone fraction is actually the real water fraction). However, apart from this ambiguity effect, the measured acetone fractions do agree closely with true acetone fractions. The agreement is closest at low flip angles (due to minimization of T1-bias), while there is some rightward skew of the plot at higher flip angles, due to increasing T1-bias.

![Figure 4-6](image.png)

**Conclusion**

Magnitude-based CSE-MRI offers good image quality and enables accurate estimates of acetone fraction (a surrogate for fat fraction) in an acetone-water fraction. The signal models here provided an accurate description of the signal evolution in an acetone-water phantom. However, fat-water ambiguity is unresolved when using magnitude imaging alone, and a separate method is required to address this problem.
The concentration-dependent variation in the size of the chemical shift between water and acetone is an interesting finding but might limit the practicality of using acetone-water phantoms as a validation or quality assurance tool.

As previously described (218), the use of low flip angles minimizes T1-bias and thus is likely to improve the accuracy of FF quantification.

4.3.2.3 ‘Berglund + magnitude’

One approach to solving the ambiguity between water-dominant and fat-dominant tissues, whilst still enabling R2* calculation, is to use a simple analytical CSE-MRI method to generate initial ‘guesses’ for the fat and water signals, followed by a magnitude-based fit to refine these guesses. In this experiment, this possibility was investigated using a three-point method recently developed by Berglund et al. (198), in conjunction with the magnitude-based approach described above.

Methods

The same two volunteers’ CSE-MRI data were used for this experiment. An outline of the processing steps performed is shown in Figure 4-7.

The raw complex data were passed to the algorithm described by Berglund, which used data from the first three echo times to generate fat and water images (198). The fat and water signal intensities from these images were then used as initial guesses for the magnitude fit (specifically, to the start points for the magnitude fit were defined equal to the signal intensities on the fat and water images generated by the Berglund method). The magnitude fit (which operated on all six data points) returned fat and water images and R2* maps. FF maps were calculated using the fat and water magnitude images and assessed for artifact severity using the qualitative scale described previously.
The MATLAB code from the ISMRM fat-water toolbox (186) was used to implement the Berglund algorithm. The magnitude fitting was performed using the in-house implementation of the method described by Bydder (218) (see Appendix A, Section 9.2.2).

**Results and Discussion**

FF maps generated using the Berglund method alone and using the ‘Berglund + magnitude’ method are shown in Figure 4-8.

For the Berglund method alone, there were no fat-water swaps in Subject 1 for either method. For Subject 2, artifact severity was moderate (swaps in 10-25% of pixels), with fat-water swaps affecting the anterior abdominal wall musculature (the bone was unaffected). The images from Subject 2 show a gradation of FF values in the anteroposterior direction (FF values are lower anteriorly), which contributes to
the swapped subcutaneous fat. This gradation may be due to phase errors related to the implementation of the readout gradients, which are known to be a potential cause of fat-water swaps (275).

Figure 4-8 - Examples of images from Berglund and 'Berglund + magnitude' methods.
For the Berglund + magnitude method, there were again no fat-water swaps in Subject 1 and moderate swaps for Subject 2. Unsurprisingly, the swaps occurred in the same locations for the two methods (where the first guess was incorrect, the final FF map was also incorrect). The Berglund + magnitude method did largely eliminate the observed anteroposterior gradation in FF values; however, the anterior-most subcutaneous fat pixels (where the effect of this gradient is most severe) were swapped. The results suggest that, as long as the Berglund method correctly defines the voxel as fat-dominant (i.e. FF above 50%) or water-dominant (i.e. FF below 50%), an accurate estimate is achieved using the Berglund + magnitude method.

The R2* maps showed good image quality, with no substantial artifacts. Note that bone shows substantially higher R2* values than surrounding soft tissue, likely due to a high R2’ due to the presence of bone trabeculae.

Conclusion

The combination of the Berglund method and the magnitude fit enables resolution of fat-water ambiguity, whilst also enabling the generation of R2* maps. This method is feasible in volunteers and could be used for subsequent research. Since this method implements R2* correction and uses a multispectral model for fat composition, the measurement can be regarded as a proton density fat fraction (PDFF).

4.3.2.4 Commercial CSE-MRI Packages

Recently, several manufacturers have developed packages for quantitative CSE-MRI in the liver, which could potentially be applied to bone imaging. At our institution, a Philips package known as mDixon Quant became available during the course of this project. Therefore, this algorithm was used ‘off-label’ and evaluated as a potential method for FF quantification in bone.

This approach has several advantages:

1. Using a manufacturer-supplied package allows for in-line processing, which saves time and simplifies the analysis.
2. The software has been developed by the manufacturer for use with their own hardware; by contrast, other methods may have been developed with different coils and sequences. This may improve image quality and accuracy.
However, disadvantages include:

1. The manufacturers’ software is proprietary, and consequently full details of the processing algorithm are not available. This means that improving the algorithm, or overcoming any problems with it, is more difficult.
2. The cost of the software is a barrier to implementation at other sites.

**Methods**

*mDixon Quant method*

In addition to the raw complex data acquired as described above, fat, water, fat fraction and R2* images were acquired from the two volunteers using a vendor-supplied gradient echo sequence (Philips mDixon Quant; Philips Healthcare, Andover, Massachusetts, USA; software version 5.1.7), which acquires six evenly-spaced echoes (first echo time 1.17 ms, echo spacing 1.6 ms) using bipolar readout gradients (39–41). PDFF and R2* maps were generated inline, assuming a 10-peak model of human adipose tissue, as previously described (40,42). Other sequences parameters included: time to repetition (TR) 25 ms, flip angle 3°, matrix size 320 x 320, pixel spacing 1.76 x 1.76 mm, and bandwidth 1894 Hz/Px.

**Assessment of image quality**

Images were qualitatively assessed (as above) to determine the severity of fat-water swaps.

**Comparison of accuracy with ‘Berglund + magnitude’**

A fat-water phantom consisting of twelve 50ml centrifuge tubes with varying true fat volume percentages was constructed based on phantoms described previously (206,235), with fat percentages chosen to reflect the range of fat fraction values observed in both normal and pathological bone marrow. True fat volume percentages in the phantom were 0, 10, 20, 30, 40, 45, 50, 55, 60, 65, 70 and 100%. Peanut oil was used as a surrogate for human fat since its NMR spectrum is similar to that of human adipose tissue (199,276), and comprises approximately 9% palmitic, 4% stearic, 55% oleic and 27% linoleic acids (277). Fat volume percentages were used as reference fat fractions, as previously described (226). For each tube, the appropriate volume of peanut oil was dispensed by weight, assuming the density of peanut oil (0.916g/cm³). Sodium dodecyl sulphate (SDS) (surfactant, Sigma Aldrich, St Louis, USA) was added to the peanut oil and gently mixed to form an initial
emulsion, ensuring a final SDS concentration in each phantom of 28mM. Appropriate volumes of 3.0% w/v agar solution preheated to 90°C were then added to each tube, and each tube was mixed by gentle inversion for approximately two minutes. The tubes cooled at room temperature and all formed a solid gel (with the exception of the 100% fat fraction tube, which contained pure oil).

The phantom was then scanned using both the mDixon Quant method (bipolar readout, TE1 1.44ms, ΔTE, 1.29ms, TR 25ms, resolution 0.52 x 0.52 x 1mm, bandwidth 1894Hz/pixel, flip angle 3°) and using the Berglund + magnitude method (monopolar readout, TE1 1.27ms, ΔTE, 1.97ms, resolution 0.93 x 0.93 x 1.5mm, bandwidth 1894 Hz/pixel, flip angle 3°).

To assess the accuracy of fat fraction measurements, linear regression was performed between known and measured fat fractions for the two methods. Two-sided t-tests were used to determine whether there were statistically significant differences between obtained slope values and 1.0, and obtained intercept values and 0.0 (α=0.05). In addition, analysis of covariance (ANCOVA) was used to determine whether significant differences existed between the two methods.

![Fat fraction and R2* maps generated using Philips mDixon Quant.](image)

Figure 4-9 – Fat fraction and R2* maps generated using Philips mDixon Quant.
Results and Discussion

FF and R2* maps generated using the mDixon Quant method are shown for Subjects 1 and 2 in Figure 4-9. Image quality was excellent - there were no visible fat-water swaps in either case.

Comparison of accuracy with 'Berglund + magnitude'

PDFF measurements from the fat-water phantom for the mDixon Quant and the Berglund + magnitude methods are shown in Figure 4-10. R² values were 0.99 for mDixon Quant and 0.97 for Berglund + magnitude. Regression values (mean ± standard error) were as follows: mDixon Quant: slope 1.027 ± 0.027, intercept 0.36 ± 1.45; Berglund + magnitude: slope 1.048 ± 0.055, intercept -2.7 ± 2.9. Neither slope value was significantly different from 1 (p=0.35 and 0.40 respectively) and there was no significant difference between the two slope or intercept values (P=0.74 and 0.35 respectively).

Conclusion

The mDixon Quant algorithm produced excellent image quality and accurate fat fraction measurements in the fat-water phantom. Both the mDixon Quant and the Berglund + magnitude methods produced accurate PDFF measurements in the fat-water phantom, with little difference between the two methods.
4.4 Conclusion

Of the methods investigated, two have been identified which have the desired attributes and can be implemented on the 3T system used for this work.

The first of these is the 'Berglund + magnitude' method, which requires the acquisition of raw complex data from the scanner and offline processing. This method uses freely-available code (which can therefore be interrogated and adjusted by the user) and could theoretically be implemented using data from most modern scanners, with no extra software requirements or additional cost. This is particularly attractive when considering implementation at multiple sites. This approach could also be implemented using the manufacturer’s own two- or three-point Dixon images (rather than the Berglund images) as initial guesses for the magnitude fit, offering some flexibility. The fitting can be adjusted if needed, for example to enable separate R2* terms for fat and water. Finally, the field map estimate from the Berglund method could also be useful for other applications, such as quantitative susceptibility mapping (see Chapter 7).

The second of these is the Philips mDixon Quant method. Although this method is designed for liver imaging, there is no theoretical reason why it cannot be applied to the bone. Preliminary data suggest that the mDixon Quant method results in good image quality and accuracy. However, the software for this method is expensive and is not widely available at other sites. Furthermore, full details of the processing algorithm are not freely available.

The relative strengths and weaknesses of the two algorithms have dictated where these should be implemented for subsequent studies. For small studies on our own scanner, the mDixon Quant method has the advantage of being quick and simple, thereby reducing processing time, and probably offers a slight advantage in terms of image quality. However, for implementation at other sites where quantitative CSE-MRI packages are not available, and to achieve a standardized protocol for multisite studies, the Berglund + magnitude method may be a good alternative.
5 PDFF and R2* as Inflammatory Biomarkers: Proof-of-Principle

5.1 Introduction

The previous Chapter outlined the rationale for using CSE-MRI to image bone inflammation, and has identified specific CSE-MRI algorithms which can be used for this purpose. The work presented in this Chapter aimed to technically validate CSE-MRI in terms of accuracy and precision (according to the metrology framework described by Sullivan (1)), and to demonstrate the relationship of PDFF and R2* measurements with active inflammation and structural damage (particularly fat metaplasia).

One important issue is that PDFF measurements can be confounded by variations in R2*, and, conversely, R2* measurements can be confounded by variations in PDFF. Therefore, to assess the relationship between BMD and R2* in the presence of fat, and between known fat fractions and PDFF measurements in the presence of bone, a series of phantoms were constructed containing varying proportions of fat, water and trabecular bone, and imaged using CSE-MRI. *In vivo*, CSE-MRI was performed in the sacroiliac joints of patients with spondyloarthritis and differences in PDFF and R2* between areas of active inflammation, fat metaplasia and normal marrow were assessed.

5.2 Methods

This study received ethical approval from the Queen Square Research Ethics Committee, London (Research Ethics Committee reference 15/LO/1475). All participants gave written informed consent prior to study entry.

5.2.1 Fat-Water and Fat-Water-Bone Phantoms

First, a fat-water phantom consisting of true fat volume percentages of 0, 10, 20, 30, 40, 45, 50, 55, 60, 65, 70 and 100% was constructed as described in Section 4.3.2.4. Photographs of the 0%, 40% and 70% FF tubes from this phantom are shown in Figure 5-1a; microscopic images of a drop of the fresh emulsion (Zeiss LSM 710 confocal microscope, Oberkochen, Germany) are shown in Figure 5-1b.
Subsequently, a fat-water-bone phantom consisting of fat-water mixtures with the addition of granules of decellularised bovine bone matrix (NuOss granules, particle size 250-1000µm, Henry Schein, London) was constructed. The bovine bone matrix granules are physically and chemically similar to the mineral matrix of human bone, and retain their natural trabecular structure. The mass of bone matrix in the tubes was chosen to reflect the range of trabecular bone mineral density values occurring in healthy subjects (278) and disease states (BMD values were 0, 50, 100 and 150mg/cm³). Phantoms were constructed in 5ml scintillation vials. To ensure homogeneity, the bone granules were first thoroughly mixed with oil, before sequential addition of SDS and heated agar. Final mixtures of fat, bone and agar were then continually agitated with a high speed touch-mixer (Vortex-Genie 2, Cole Palmer, London, UK) until solid. All tubes formed a solid gel. The volume of the bone granules was much smaller than the volume of fluid in each vial (i.e. unmeasurable); the volume of the bone granules was therefore neglected in calculations of BMD.

The arrangement of the fat-water-bone phantom is shown in Figure 5-1c; microscopic images of fat-water emulsions containing trabecular bone are shown in Figure 5-1d.

5.2.2 Patients and Volunteers

Subjects were prospectively recruited by rheumatologists working in a specialist tertiary rheumatology clinic. Patients aged between 12 and 30 years were included in the study if they had a known diagnosis of sacroiliitis (diagnosed previously on magnetic resonance imaging) or were suspected of having sacroiliitis after assessment by a rheumatologist. All patients with sacroiliitis fulfilled diagnostic criteria for enthesitis-related arthritis (with axial involvement) or non-radiographic axial spondyloarthritis (6,7,69). Patients with suspected sacroiliitis at the time of the scan were subsequently treated as controls if the clinical MRI scan was normal. In addition, eight healthy volunteers aged between 18 and 30 years had repeat MRI scans of the sacroiliac joints one month apart to assess measurement repeatability.
Figure 5-1 - Fat-water (a,b) and fat-water-bone (c,d) phantoms. a: 0%, 30% and 70% FF vials are shown lying flat to demonstrate solidity. b: Microscopic images of 30% FF tube shows the arrangement of fat and water in the emulsion (the fat globules are circular and outlined by dark lines). c: Photograph of the fat-water-bone phantom, showing fat content varying by row and BMD varying by column. d: Microscopic images of trabecular bone granules in dilute emulsion (approximately 10% FF); a number of fat droplets are visible coating the surface of the trabeculae.
Figure 5-2 - Delineation of oedema and fat metaplasia. Areas of bone marrow oedema and fat metaplasia were identified on T2-weighted STIR (a) and T1-weighted (d) spin echo images respectively. Freehand ROIs were placed on these images (b, e) and directly transferred to the corresponding anatomical location on PDFF maps (c, f) using the sacroiliac joints and sacral foramina as anatomical landmarks.
5.2.3 Imaging Sequence and Reconstruction

CSE-MRI of phantoms and all subjects was performed on a wide-bore 3.0T clinical system (Ingenia, Philips, Amsterdam, Netherlands) with integrated posterior and anterior surface coils using a three-dimensional spoiled gradient echo recalled (SPGR) acquisition. Images of patients and volunteers were acquired in a tilted coronal plane parallel to the long axis of the sacrum, and 40 slices were acquired with a slice thickness of 2mm. Imaging was performed using a vendor-supplied gradient echo sequence (Philips mDixon Quant, software version 5.1.7) which acquires six evenly-spaced echoes (first echo time 1.17ms, echo spacing 1.6ms) using bipolar readout gradients (270, 279, 280). PDFF and R2* maps were generated in-line, assuming a 10-peak model of human adipose tissue as previously described (201,270). Other sequence parameters included: repetition time 25ms, flip angle 3°, repetition time 25ms, matrix size 320 x 320, pixel spacing 1.76mm x 1.76mm, and bandwidth 1894 Hz/Px. Parallel imaging was not used. Additionally, all patients underwent a standard clinical MRI scan on the same day on a 1.5T system (Avanto, Siemens, Berlin, Germany) with coronal T1-weighted and T2-weighted STIR sequences (119) which were acquired in the same tilted coronal plane as the CSE scans.

5.2.4 Statistical Analysis

For the phantom experiments, statistical analyses were based on those used in previous phantom studies using fat-water phantoms (199). PDFF and R2* measurements were performed using circular regions-of-interest (ROIs) placed on axial slices through the phantom tubes, taking care to avoid the edges of the emulsions. ROIs were placed in identical locations on PDFF and R2* maps. To assess the accuracy of fat fraction measurements, linear regression was performed between known and measured fat fractions and two-sided t-tests were used to determine whether there were statistically significant differences between obtained slope values and 1.0, and obtained intercept values and 0.0 (α=0.05). To assess the relationship between BMD and R2*, linear regression was performed between measured R2* values and known BMD values, and analysis of covariance (ANCOVA) was used to determine whether obtained slope and intercept values were significantly different in vials containing fat compared to 0% fat fraction vials.
For the patient study, areas of bone marrow oedema and/or fat metaplasia in patients with sacroiliitis were identified on T₂-weighted STIR and T₁-weighted images respectively, and freehand regions-of-interest (ROIs) were placed on those areas, taking care to avoid normal bone, by a radiology registrar (TB) with four years of experience in musculoskeletal MRI (see Figure 5-2). A maximum of four ROIs for bone marrow oedema and four ROIs for fat metaplasia were used in each subject. Using the synovial and ligamentous portions of the sacroiliac joints and sacral foraminae as fixed anatomical landmarks, ROIs were then transferred manually onto PDFF maps at the same anatomical location, and mean PDFF and R₂* values were recorded for each ROI. In controls with mechanical back pain, freehand ROIs were placed on the largest possible areas of subchondral bone marrow on both the sacral and ilial sides of the joint, for both the left and right sacroiliac joints on a single slice. To determine whether there were statistically significant differences in PDFF and R₂* between areas of normal bone marrow, oedema and fat metaplasia, data were analysed using a multilevel mixed-effects linear regression model on Stata software (version 14.1, StataCorp, College Station, TX) which accounted for repeated observations in individual patients. ‘Tissue type’ (i.e. normal marrow, oedema or fat metaplasia) was used as a predictor variable, PDFF or R₂* were used as outcome variables, and subject number was included as a grouping variable.

To assess the repeatability of PDFF and R₂* measurements, four ROIs (one ROI for each side of the joint, for both left and right sacroiliac joints) were placed on the subchondral bone using a single representative slice through the sacroiliac joints. ROI data were analysed using Bland-Altman 95% limits of agreement with adjustment of the limits to account for multiple observations for each individual (281), and also using the intraclass correlation coefficient.

5.3 Results

5.3.1 Fat-Water and Fat-Water-Bone Phantoms

PDFF maps from the fat-water phantom are shown in Figure 5-3a; PDFF and R₂* maps from the fat-water-bone phantom are shown in Figure 5-3b,c.

Figure 5-4a,b show the relationship between reference fat fraction and measured PDFF values in the fat-water and fat-water-bone phantoms. In the absence of bone
(Figure 5-4a), measured PDFF values agreed closely with reference fat fraction values \((\text{slope } = 1.03 \pm 0.027, \text{intercept } = 0.0036 \pm 0.014)\). Despite the presence of trabecular bone, measured PDFF values also agreed closely with reference fat fraction values in the fat-water-bone phantom (Figure 5-4b): results of the linear regression analysis are given in Table 7.

<table>
<thead>
<tr>
<th>BMD</th>
<th>0mg/cm³</th>
<th>50mg/cm³</th>
<th>100mg/cm³</th>
<th>150mg/cm³</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R^2)</td>
<td>0.989</td>
<td>0.998</td>
<td>0.991</td>
<td>0.998</td>
</tr>
<tr>
<td>Slope</td>
<td>1.03 ± 0.062</td>
<td>1.00 ± 0.026</td>
<td>0.99 ± 0.053</td>
<td>1.01 ± 0.029</td>
</tr>
<tr>
<td>(P = 0.59)</td>
<td>(P = 0.90)</td>
<td>(P = 0.85)</td>
<td>(P = 0.71)</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.018 ± 0.025</td>
<td>0.010 ± 0.011</td>
<td>0.0090 ± 0.021</td>
<td>-0.008 ± 0.012</td>
</tr>
<tr>
<td>(P = 0.47)</td>
<td>(P = 0.34)</td>
<td>(P = 0.68)</td>
<td>(P = 0.49)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reference fat fraction</th>
<th>0%</th>
<th>20%</th>
<th>40%</th>
<th>50%</th>
<th>60%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R^2)</td>
<td>0.956</td>
<td>0.985</td>
<td>0.841</td>
<td>0.941</td>
<td>0.997</td>
</tr>
<tr>
<td>Slope</td>
<td>0.00056 ± 8.5x10^{-5}</td>
<td>0.00090 ± 7.7x10^{-5}</td>
<td>0.00066 ± 0.00020</td>
<td>0.00080 ± 0.00014</td>
<td>0.0016 ± 6.3x10^{-5}</td>
</tr>
<tr>
<td>(P = 0.39)</td>
<td>(P = 0.99)</td>
<td>(P = 0.67)</td>
<td>(P = 0.002)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.023 ± 0.055 ± 0.088 ± 0.019</td>
<td>0.085 ± 0.013</td>
<td>0.035 ± 0.0059</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0079</td>
<td>0.0072</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(P = 0.27)</td>
<td>(P = 0.014)</td>
<td>(P = 0.026)</td>
<td>(P = 0.88)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7 - Results of linear regression analysis (slope and intercept values are given as estimate ± standard error). (a) Measured PDFF values agreed closely with known fat fractions for all BMD values in the phantom. (b) Measured R2* values were linearly related to BMD (all slope values were significantly greater than 0), although the slope varied significantly depending on the fat fraction. \(P\)-values relate to comparison with the 0% fat fraction vial using ANCOVA.
Chapter 5 - PDFF and R2* as Inflammatory Biomarkers: Proof-of-Principle

Figure 5-3 - Fat–water and fat–water–bone phantoms. (a) PDFF maps from the fat–water phantom. Labels indicate reference FF values, which vary between 0 and 100%. (b and c) PDFF and R2* maps from the fat–water–bone phantom. Reference FF values vary by row, whereas bone mineral density values vary by column.

Figure 5-4 - Relationship of PDFF and R2* with reference FF and BMD. (a) PDFF measurements agree closely with reference FFs in the absence of bone. (b) Increases in bone density have minimal impact on PDFF measurements over a physiological BMD range. (c) R2* measurements are linearly related to BMD, although the slope varies depending on fat content. (d) Variations in FF also influence the measured R2* for a given BMD.
Figure 5-4c shows the relationship between trabecular BMD and measured R2* values, and results of the linear regression analysis are given in Table 7b. In the absence of fat, there was a significant linear relationship between trabecular BMD and R2*. Linear relationships were also observed in vials containing fat, although slope values were increased compared to the 0% fat fraction vial. Comparison of slope values between vials containing fat (i.e. between 20%, 40%, 50% and 60% fat fraction vials) revealed no significant differences between fat fractions of 20%, 40% and 50%, although slope values were significantly increased at 60% fat fraction. Figure 5-4d shows the relationship between R2* and reference fat fraction. Although the nature of this relationship was complex, increases in fat fraction at higher bone concentrations tended to increase R2* measurements.

5.3.2 Patients and Volunteers

For the patient study, 18 subjects (12 males and 6 females) were prospectively recruited with a mean age of 17y5m ± 2y11m (mean ± SD). Of these, eight subjects (six males and two females, with mean age 18y2m ± 3y4m) had evidence of active or chronic sacroiliitis using conventional MRI. The remaining ten subjects (6 males and 4 females, with mean age 16y11m ± 2y6m) had normal conventional MRI scans and no clinical or biochemical evidence of sacroiliitis, and were therefore treated as controls. Images from all subjects were of high quality; there were minor artifacts at the edges of the field of view in some subjects but these did not overlap the pelvis or sacroiliac joints.

PDFF and R2* measurements are compared in normal marrow, marrow oedema and fat metaplasia in Figure 5-5. Example PDFF and R2* maps in patients with active and chronic sacroiliitis are shown in Figure 5-6. Median (IQR) PDFF values were 46.8% (39.5 - 52.5%) in normal bone marrow, 26.9% (21.5-28.2%) in bone marrow oedema and 82.3% (75.1 - 87.8%) in fat metaplasia. Compared to normal marrow, PDFF measurements were significantly lower in areas of oedema (P=0.047) and were significantly higher in areas of fat metaplasia (P=0.000). Median (IQR) R2* values were 0.126ms⁻¹ (0.117 – 0.149ms⁻¹) in normal marrow, 0.112ms⁻¹ (0.102 - 0.136ms⁻¹) in bone marrow oedema and 0.083ms⁻¹ (0.056 - 0.100 ms⁻¹) in fat metaplasia. R2* was significantly lower in areas of fat metaplasia (P=0.005) but there was no significant difference in R2* between normal bone and bone marrow oedema (P=0.461).
Figure 5-5 - PDFF and R2* values in areas of normal bone marrow, bone marrow edema, and fat metaplasia. (a) PDFF values were significantly increased in areas of fat metaplasia, and significantly reduced in areas of bone marrow edema compared to normal subchondral marrow. (b) R2* values were significantly reduced in areas of fat metaplasia but not significantly altered in areas of edema. (c) PDFF and R2* measurements are shown on a scatterplot; areas of fat metaplasia exhibit both increases in PDFF and reductions in R2*.

Figure 5-6 - PDFF maps (a, b) and R2* maps (c, d) demonstrating bone marrow edema and fat metaplasia in patients with sacroilitis. Bone marrow edema (a, c; solid red arrow) causes a decrease in PDFF but no change in R2*. Fat metaplasia (b, d; dashed red arrow) causes an increase in PDFF and a reduction in R2*. The reduction in R2* may indicate a loss of bone mineral density.
5.3.3 Repeatability

For repeatability assessment, eight healthy volunteers (seven males and one female, mean age 31y0m ± 3y5m) were recruited and scanned on two occasions one month apart. After adjusting the limits to account for multiple observations per individual, Bland Altman 95% limits of agreement were -1.1 ± 9.9\% for PDFF and -0.0005 ± 0.075ms\(^{-1}\) for R2* (Figure 5-7). Intraclass correlation coefficients were 0.87 for PDFF and 0.98 for R2*.

Figure 5-7 - Scatterplots and Bland-Altman plots for PDFF and R2* repeated measurements.

5.4 Discussion

Here, we present a new approach to the assessment of active and chronic joint inflammation in spondyloarthritis, using CSE-MRI. We show that PDFF measurements accurately reflect changes in bone marrow composition in areas of oedema and fat metaplasia, which can be viewed as active inflammatory and
structural lesions respectively. Additionally, we show that R2* measurements are reduced in areas of fat metaplasia, which may be due to a reduction in bone mineral density. Inflammation and structural damage are hallmarks of spondyloarthritis and CSE-MRI could have a multitude of clinical and research applications in this disease, including monitoring inflammatory activity and structural damage over time. Although we focused on spondyloarthritis in this study, it is anticipated that quantitative CSE-MRI could find utility in other diseases affecting the bone marrow, such as multiple myeloma or osteomyelitis.

In our fat-water and fat-water-bone phantom experiments, PDFF measurements agreed closely with reference fat fraction values over a physiologically-relevant range, and were essentially unaffected by variations in BMD. In patients, PDFF measurements were significantly increased in areas of bone marrow oedema, and significantly reduced in areas of fat metaplasia, suggesting that PDFF is altered at both actively and chronically inflamed sites. Decreases in PDFF in oedematous bone is probably due to the presence of an inflammatory exudate (282), whilst increases in marrow fat content may be due to increases in the proportion or size of adipocytes in the marrow. Given that fat metaplasia is known to be preceded by inflammation (115,155), our data imply that the bone marrow undergoes biphasic changes in fat content with inflammation, which can be directly measured using CSE-MRI.

Our phantom experiments also indicated a positive association between R2* and BMD (222,223). A number of authors have previously explored the use of R2* or R2' as markers of osteoporosis, and these parameters have been found to parallel apparent density measured by dual-energy X-ray absorptiometry (DEXA) or quantitative computed tomography (QCT) (159,229–232). However, our results indicate that R2* measurements are also influenced by PDFF and cannot be regarded as a 'pure' marker of BMD when fat content also varies. Importantly, R2* values are also influenced by trabecular geometry (283), by the assumption of a single 'system' R2* decay term for the marrow (218), and by the arrangement of the fat and water compartments (284). For example, in patients with muscular dystrophy, the highest R2* values occur in muscles where fat and normal muscle are interdigitated, whereas R2* measurements in patients with heavily fat-infiltrated muscles are comparable to those seen in normal subjects (284). Nonetheless, our results are somewhat different to those observed in these patients with muscular dystrophy in that areas of bone marrow oedema had lower fat fraction values than
normal marrow but did not undergo a significant reduction in R2*. Conversely, there was a significant reduction in R2* in areas of fat metaplasia. Given the known relationship between R2* and BMD, this result may be due to a reduction in trabecular bone mineral density. Previous studies also support this suggestion: for example, Kühn et al. showed that R2* measurements could discriminate normal and osteoporotic patients with reasonably high levels of sensitivity and specificity, despite concomitant variations in fat content in osteoporotic individuals (159).

From a biological perspective, reductions in BMD in areas of fat metaplasia might be expected because inflammatory cytokines such as tumour necrosis factor alpha (TNFα) are known to increase osteoclastic activity (59,61,62). Additionally, increases in fat content and reduced bone mineral density in areas of fat metaplasia could occur through a common mechanism; previous studies suggest differentiation of marrow stem cells into adipocytes occurs at the expense of osteoblasts (285) and it may be that this balance has been perturbed by previous inflammation.

There is an extensive literature on the use of qualitative scoring systems to evaluate inflammation of the sacroiliac joints (119,155,163), but there have been relatively few studies using quantitative MRI for this purpose. Several small studies have demonstrated that diffusion-weighted MRI can be used to quantify inflammation in both adults and children (243,247,252,286), but problems with the reproducibility of this technique across different scanning platforms and sites may limit its widespread application (287,288). There has been very little previous research using CSE-MRI to evaluate marrow inflammation. A conference abstract by Lee et al. briefly describes the use of three-point Dixon MRI to measure changes in fat fraction at vertebral corners in spondyloarthritis (289), but this work did not utilise R2* correction or measurement, did not differentiate between fat fraction changes occurring in acute and chronic inflammation, and did not involve specific evaluation of the sacroiliac joints.

As a method for quantifying joint inflammation, the proposed method has several advantages over existing approaches. Firstly, the repeatability of PDFF and R2* measurements was excellent, and differences between repeated measurements were much smaller than differences between normal bone, oedema and fat metaplasia. ICC values were superior to previously-quoted values for visual scoring (155). Secondly, the volumetric nature of the acquisition means that images can be reformatted in
multiple planes without the need for repeated acquisitions; using conventional MRI, repeated images are typically acquired in axial, coronal and sometimes also sagittal planes (119). CSE-MRI may therefore help to reduce scan time and could easily be applied to whole body imaging. Whole body MRI can be used to identify occult inflammation and facilitate early diagnosis (290,291), and the method proposed here is likely to be much faster than existing spin echo-based approaches. Even greater acceleration could be achieved using compressed sensing and parallel imaging techniques, as recently described (292). CSE-MRI might also be used to generate whole body R2* maps, which could be used to monitor structural damage and progression throughout the spine.

The fat-water-bone phantoms described here provide a potential framework for refining PDFF and R2* measurements, and evaluating the accuracy of individual CSE-MRI methods in the bone marrow. Although previously authors have described phantoms in which fat-water mixtures are used to fill cadaveric human bone (226), the approach described here allows for direct manipulation of trabecular bone mineral density using standardized bone granules and could easily be replicated at other sites. Fat-water-bone phantoms may be useful because differences in hardware, sequences and post-processing techniques between imaging platforms could all potentially influence PDFF and R2* measurements. The optimal method is likely to depend on the specific sequence, imaging platform, image reconstruction software and anatomical region. Ideally, it would be possible to compare PDFF and R2* measurements across multiple sites, and further research is required to establish the optimal imaging method on different platforms and to establish the reproducibility of these measurements.

A limitation of this study is that the analyses used here assumed a multipeak fat spectrum based on measurements from human adipose tissue rather than the bone marrow. However, previous studies suggest that spectra from bone marrow and human adipose tissue are extremely similar (201). Furthermore, a recent study by Wang et al. compared a number of spectral models of fat and found only small differences in PDFF measurements between multipeak models, with no specific choice of multipeak model superior to the other models tested (293). Additionally, the reference fat fraction values described in this study do not represent true gold standard PDFF measurements, which would ideally be performed using spectroscopy. Nonetheless, reference fat fractions and true PDFF values are likely to

164
agree relatively closely, and the conclusions drawn here are not contingent on precise assessment of accuracy in phantoms. It is unclear whether the fat spectrum in areas of fat metaplasia is altered compared to normal marrow: accurate characterization of the fat spectrum in areas of fat metaplasia (for example using spectroscopy) would help to clarify this issue. Accurate spectral fat modelling may also help to reduce confounding effects on R2* measurements (294). Finally, variations in temperature alter the magnitude of chemical shift between fat and water (295), which was not accounted for in this study. This may have a small impact on the phantom experiments, which were performed at room temperature rather than body temperature.

The patients included in this study represent a relatively young spondyloarthritis cohort – most subjects were in their teens. Although imaging is arguably of particular importance in this early-onset group (67), changes in skeletal maturity may influence PDFF and R2* measurements and therefore make quantification of inflammation more difficult. Younger patients may have more cellular (i.e. less fatty) marrow and unossified juxta-articular bone, which may mimic bone marrow oedema (263). Maturity-related changes in marrow composition could potentially be adjusted for using background marrow as a reference; it is also possible that areas of immature bone could be differentiated based on their R2* measurements, which would be expected to be lower than those in fully ossified, mature bone. Ultimately, it would be desirable to develop a technique which could be used to quantify inflammation across the full age spectrum.

A specific limitation of the fat-water-bone phantom is that, at high fat fraction and BMD values, the phantoms were not perfectly homogenous. PDFF values appeared slightly higher at the edges of the emulsion, which may be due to some separation of the oil from the gel. The impact on absolute PDFF measurements for each vial appeared to be minimal, although it is likely that this inhomogeneity had some influence on the measured R2* values, and may have contributed to the complex dependence of R2* on fat fraction. Using either solid animal fats or thick liquid phantoms, for example containing carrageenan (235), may enable construction of more homogenous phantoms with high fat fractions. In the future, it would be useful to directly evaluate the relationship between BMD and R2* in vivo using DEXA or QCT, although this approach would be subject to ethical constraints given the use of ionizing radiation in a cohort of paediatric and young adult patients.
It should be noted that the bone granules used in this study contain only the mineral content of trabecular bone, meaning that the organic matrix is missing and any effect of collagen on the measured PDFF will not be present. Experiments with ex-vivo human cortical bone samples demonstrate that collagen-bound water makes up a significant fraction (approximately one third) of the water content of trabecular bone (296). At the echo-times used for the CSE-MRI in this work, the signal from this bound-water fraction will therefore be absent due to its very short T2*.

Karampinos et al. have demonstrated that this can result in overestimation of PDFF in comparison to spectroscopic methods that account for short T2* water species (224). In the future, using more sophisticated phantoms including collagen-bound water may help to overcome this bias.

5.5 Conclusion

In conclusion, we have shown that PDFF measurements reflect changes in marrow composition in areas of bone marrow oedema and fat metaplasia, and could therefore be used to monitor both active inflammation and structural damage in spondyloarthritis. R2* measurements may provide additional information about marrow structure, but are also influenced by fat content. The reduction in R2* in areas of fat metaplasia is a new finding, and may be due to loss of trabecular bone mineral density. Furthermore, we have described a series of fat-water and fat-water-bone phantoms which provide a potential framework for refining PDFF and R2* measurements, and evaluating the accuracy of individual CSE-MRI methods in the bone marrow.

5.6 Acknowledgements

This work was published in Magnetic Resonance in Medicine, with the following reference: Bray TJP, Bainbridge A, Punwani S, Ioannou Y, Hall-Craggs MA. Simultaneous Quantification of Bone Edema/Adiposity and Structure in Inflamed Bone using Chemical-Shift Encoded MRI in Spondyloarthritis. *Magnetic Resonance in Medicine 2018 Feb;79(2):1031-1042*
6  The BEACH Tool; Comparison of ADC and PDFF as Inflammatory Biomarkers in a Prospective Study

6.1  Introduction

In the previous Chapter, it was shown that PDFF and R2* measurements are altered in individual inflammatory lesions, suggesting a potential role in measuring active and chronic inflammation. However, in order to develop PDFF and R2* as clinically-useful biomarkers, it is necessary to derive an overall measure of sacroiliac joint inflammation in each patient, which is not restricted to specific areas of inflammation identified by the reader.

The linear ROI tool described in Chapter 3 can be used for this purpose, but has several limitations including the potential for sampling error and the inclusion of the sacroiliac joint space (which is not ideal when the SPARCC visual scoring system, which only assesses subchondral bone, is used as a reference standard). An additional problem is that the linear ROI tool measures mean ADC/nADC measurements, which may inadequately capture the mixture of active and chronic inflammatory changes which are common in patients with sacroiliitis. Since ADC values are increased in inflammation and likely to be reduced in fat metaplasia (due to the contribution of unsuppressed fat to the ADC measurement), patients with both active and chronic inflammation may actually have normal mean ADC values. Similarly, active inflammation causes a reduction in PDFF whilst fat metaplasia causes an increase in PDFF, meaning that mean PDFF could be normal in patients with mixed inflammation. One approach to resolving this issue is to use histographic measurements such as percentile values, which can potentially provide a more faithful representation of pathology in tissues where there are multiple pathological processes occurring simultaneously (297–301).

In this work, we present an improvement of the linear ROI method which, firstly, enables a more specific assessment of subchondral bone (with reduced potential for sampling error) and, secondly, measures a series of histographic parameters which may describe active and chronic inflammation more accurately than simple averages. The proposed method is partially automated in the sense that polygonal regions-of-interest (ROIs) are automatically propagated onto subchondral bone after the user
defines the joint, thereby reducing the subjectivity associated with ROI placement and potentially improving reproducibility. We aimed to validate this tool, which we have named BEACH (Bone OEdema and Adiposity quantification using ADC and Chemical shift imaging with histograms), in a prospective study of 53 patients.

6.2 Methods

This study received ethical approval from the Queen Square Research Ethics Committee, London, United Kingdom (Research Ethics Committee reference 15/LO/1475). All participants gave written informed consent prior to study entry.

6.2.1 Patient Selection

Fifty-three patients (mean age, 18 years; age range 12-23 years), of whom 31 were males (mean age, 18 years) and 22 were females (mean age, 17 years), were prospectively recruited (this 53 includes 18 patients whose data were initially reported in Chapter 5). Patients were included if they were referred for a clinically-indicated MRI scan of the sacroiliac joints either for suspected sacroiliitis or for monitoring of known sacroiliitis, as determined by adolescent and young adult rheumatologists working in a tertiary referral clinic, and were aged between 12 and 30. Patients were excluded if they had a contra-indication to MRI scanning.

6.2.2 Image Acquisition

All subjects underwent quantitative MRI scans, consisting of CSE-MRI and DWI acquisitions, and conventional MRI. Quantitative CSE-MR images were acquired on a 3T Philips Ingenia scanner (Ingenia, Philips, Amsterdam, Netherlands) using Philips mDixon Quant, as described in Section 5.2. DW images were acquired on a 1.5T Siemens Avanto scanner (Avanto, Siemens, Munich, Germany) with a standard Stejskal-Tanner sequence with echo-planar readout, and b-values of 0, 100, 300, 600 and 1000 s/mm². ADC maps were calculated in-line on vendor software, using a standard monoexponential fit.

Conventional MRI images consisted of T2-weighted STIR, T1-weighted and fat-suppressed post-contrast T1-weighted images, with parameters as described in Section 3.4.2.2. All conventional MR images were acquired in both angled coronal (parallel to the sacrum) and angled transverse (perpendicular to the sacrum) planes.
### 6.2.3 Definition of Regions-of-Interest

QIB measurements were obtained from the PDFF and ADC maps using the BEACH tool, which was implemented in MATLAB. The tool was designed to enable quantitative measurements to be obtained from subchondral bone [the same location as existing qualitative scoring systems (163)] in a simple and reproducible fashion, as follows.

The observer is prompted to define the line of the sacroiliac joint using a single series of connected straight lines – an open polygon – as shown below in Figure 6-1. Additionally, ‘anchor lines’ are used to define the angle made by the joint with the cortical surface, at both the top and bottom of the joint, enabling the shape of the polygonal ROIs to be closely matched to subchondral bone. The software then automatically generates a pair of polygonal ROIs in the subchondral bone either side of the joint, as shown in Figure 6-1. Further detail on the geometry of the generated ROIs is provided in the Supplementary Information (see Section 6.6.1).

![Figure 6-1 - Definition of polygonal ROIs on subchondral bone. The observer is asked to define the line of the sacroiliac joint and ‘anchor lines’ are added to define the angles made by the joint with the anterior and posterior cortex of the bone, thus enabling the automatically-propagated ROIs to better fit the subchondral bone.](image)
This procedure is repeated for both the left and right sacroiliac joints, on each slice, until the subchondral bone included in the imaging volume had been fully sampled. In the case of ADC maps, we included all slices where the synovial joint was visible, whereas alternate slices were used for the PDFF maps due to the much thinner slices (2mm) available from the volumetric CSE-MRI. To be consistent with the visual scoring systems used for comparison in this work, only the subchondral bone corresponding to the synovial part of the joint was defined (the bone corresponding to the ligamentous part of the joint was excluded).

For each patient, pixel values from the total volume of defined subchondral bone (i.e. from all ROIs) were analysed histographically, as described in the following section.

In this study, the ROIs were generated by two radiology registrars (NS and AD, with two and one years of experience in MR imaging respectively) who each received a training session (from TB) on how to identify the sacroiliac joint and how to use the tool prior to performing the measurements.

### 6.2.4 Histographic Analysis

For both PDFF and ADC, we measured the 10\textsuperscript{th}, 25\textsuperscript{th}, 50\textsuperscript{th}, 75\textsuperscript{th} and 90\textsuperscript{th} centiles of the distribution, referred to as PDFF\textsubscript{10}, PDFF\textsubscript{25}, PDFF\textsubscript{50}, PDFF\textsubscript{75} and PDFF\textsubscript{90} for PDFF, and as ADC\textsubscript{10}, ADC\textsubscript{25}, ADC\textsubscript{50}, ADC\textsubscript{75} and ADC\textsubscript{90} for ADC, respectively. Examples of PDFF histograms generated using the BEACH tool and corresponding percentile measurements are shown in Figure 6-2; ADC histograms and corresponding percentile measurements are shown in Figure 6-3. In addition to percentile measurements, we measured the proportion of PDFF values $p_{\text{low(PDFF)}}$ which were below a pre-defined threshold for inflammation of 28.2\%, and the proportion of values $p_{\text{high(PDFF)}}$ which were above a pre-defined threshold for fat metaplasia of 69.7\%. The thresholds were derived using data from ROIs on individual inflammatory lesions (i.e. areas of bone marrow oedema or fat metaplasia), as described in the supplementary information (Section 6.6.2). For ADC maps, we again measured the 10\textsuperscript{th}, 25\textsuperscript{th}, 50\textsuperscript{th}, 75\textsuperscript{th} and 90\textsuperscript{th} centiles of the distribution, denoted ADC\textsubscript{10}, ADC\textsubscript{25}, ADC\textsubscript{50}, ADC\textsubscript{75} and ADC\textsubscript{90}. We also measured the proportion of values $p_{\text{high(ADC)}}$ above a pre-defined threshold for inflammation of $957\text{mm}^2/\text{s}$. This threshold was derived using a separate dataset of patients with spondyloarthritis, as described in the supplementary information (Section 6.6.2). For each quantitative score, the mean of the two observers’ measurements was used for the analysis.
Figure 6-2 - Examples of histograms generated using the BEACH tool. Conventional MR images (a,c). PDFF maps (d,e,g) and PDFF histograms (g,h,i) are shown. In the normal patient’s histogram (g), PDFF values are clustered around 50%, corresponding to normal marrow. In the patient with inflammation, a number of low-PDFF pixels have emerged in the histogram (h). In the patient with fat metaplasia, there is an upward shift in PDFF values, with a large number of high-PDFF pixels (i).
Figure 6.3 - Examples of ADC histograms in patients with sacroiliitis (a) and control patients (b). The red lines indicate the 10th, 25th, 50th, 75th and 90th percentiles of the ADC distribution.
### 6.2.5 Scoring of Conventional MRI

Each set of conventional MR images (including STIR, T1-weighted and contrast-enhanced images) was scored by two experienced musculoskeletal radiologists (KR and MHC) with ten and over 25 years of MRI experience respectively. Images were read on a dedicated research workstation, and both readers were blinded to clinical diagnosis, treatment and to the quantitative images and measurements.

For each patient, the observer was asked to assign each of the scans a qualitative score between 0 and 72 for the extent/severity of inflammation using the Spondyloarthritis Research Consortium of Canada (SPARCC) system (163). The patient was deemed to have met the ASAS definition of active inflammation if the mean inflammation score from the two readers was $\geq 2$, $(117, 302, 303)$. Similarly, structural lesions consisting of fat metaplasia, erosions and joint ankylosis were assessed using a recent adaptation of the SPARCC system (155). To enable a binary classification of the presence/absence of each of these lesions, we adopted a definition of a positive MRI for individual structural lesions (again based on the mean score from the two observers), in which patients with a $\geq 3$ fatty lesions and/or $\geq 3$ erosions were deemed structurally abnormal (these cutoffs have been shown to produce specificity values of $>95\%$ in existing spondyloarthritis cohorts) (155,304). The mean inflammatory and structural scores were also used as measures of the burden of inflammation and fat metaplasia, respectively, for the regression analysis with qMRI scores (see Statistical Analysis, below).

### 6.2.6 Clinical Scores

For each patient, symptoms were assessed using a dedicated research questionnaire designed by an adolescent and young adult rheumatologist (CF), which included the Bath Ankylosing Spondylitis Disability Index (BASDAI), Bath Ankylosing Spondylitis Functional Index (BASFI), Health Activity Questionnaire (HAQ), Work Productivity Impairment Questionnaire plus Classroom Impairment Questions (WPAI+CIQ), Patient Global Assessment, Jenkin’s Sleep scale and Margolis Pain Diagram. Additionally, all patients were clinically examined by a specialist adolescent rheumatologist at the time of scan referral to determine the presence of swollen, restricted or active joints, and of enthesitis. For brevity, a selected subset of these clinical assessment scores have been included in the results of this study.
6.2.7 Statistical analysis

Quantitative parameters derived from ADC and PDFF maps were first assessed for normality using the D’Agostino-Pearson omnibus normality test, and were then compared between patients with and without SIJ inflammation, and between patients with and without SIJ fat metaplasia, using two-sided t-tests. Receiver-operating characteristic analyses were used to assess the sensitivity and specificity with which different quantitative parameters could distinguish between patients with and without SIJ inflammation, and between patients with and without SIJ fat metaplasia. Specifically, we measured the area under the curve (AUC) for distinguishing any patient with inflammation (i.e. meeting the ASAS definition for active inflammation) from all other subjects, and also the AUC for distinguishing any patient with fat metaplasia from all other subjects. AUC values were described as follows: 0.5-0.6: fail, 0.6-0.7: poor, 0.7-0.8: fair, 0.8-0.9: good, 0.9-1: excellent.

To assess the relationship between the qualitative severity scores and the quantitative scores, linear regression was used with the qMRI measurement (specifically the best-performing parameters from the ROC analysis) as the predictor variable and the qualitative score as the outcome variable, with two-sided t-tests used to determine whether there were statistically significant differences between obtained slope values and 0.0. Linear regression was also used to examine the relationship between clinical scores (BASDAI, BASFI, CRP and ESR) and visual scoring, and between clinical scores and qMRI parameters.

Inter-observer variability was assessed using Bland-Altman 95% limits of agreement and the intraclass correlation coefficient for both the visual scores and qMRI parameters.

6.3 Results

6.3.1 Inflammatory Parameters

15/53 patients met the ASAS criteria for active inflammation, and 38/53 patients did not (these patients are referred to as inflamed and uninflamed respectively). There were more males (12) than females (3) in the inflamed group, but the difference in gender ratio between inflamed and uninflamed groups was not significant ($P = 0.065$). There was no significant difference in age between inflamed and uninflamed groups ($P = 0.43$).
Comparisons of quantitative parameters between inflamed and uninflamed SIJs are shown in Figure 6-4 (for ADC) and Figure 6-5 (for PDFF), and results of the corresponding t-tests and ROC analyses are shown in Table 8.

All of the investigated ADC-based parameters were significantly higher in inflamed SIJs than in uninflamed SIJs. The ability of these parameters to distinguish between inflamed and uninflamed SIJs was improved for parameters which specifically sampled the upper end of the ADC distribution (i.e. $\text{ADC}_{75}$, $\text{ADC}_{90}$ and $p_{\text{high}(\text{ADC})}$). The best performing parameter was $\text{ADC}_{90}$, which produced an AUC value of 0.819.

PDFF-based parameters performed poorly as measures of inflammation, and there was no significant difference between inflamed and uninflamed SIJs for any of the investigated parameters. Nonetheless there was a trend towards significance was observed for parameters which specifically sampled the lower end of the PDFF distribution ($P = 0.06$ for PDFF$_{10}$), with a corresponding increase in performance (AUC = 0.658 for PDFF$_{10}$ compared to AUC = 0.514 for the median PDFF).

---

**Figure 6-4 - ADC as an inflammatory marker.**
Chapter 6 - The BEACH Tool; Comparison of ADC and PDFF as Inflammatory Biomarkers in a Prospective Study

Figure 6-5 - PDFF as an inflammatory marker.

<table>
<thead>
<tr>
<th>Inflammation</th>
<th>Present (n=15)</th>
<th>Absent (n=38)</th>
<th>ROC AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADC_mean</td>
<td>7.58 ± 0.66</td>
<td>5.78 ± 0.22</td>
<td>( P = 0.0014^{**} )</td>
</tr>
<tr>
<td>ADC_median</td>
<td>7.47 ± 0.64</td>
<td>5.77 ± 0.22</td>
<td>( P = 0.0025^{**} )</td>
</tr>
<tr>
<td>ADC_25</td>
<td>9.64 ± 0.77</td>
<td>7.33 ± 0.19</td>
<td>( P = 0.0001^{***} )</td>
</tr>
<tr>
<td>ADC_100</td>
<td>11.72 ± 0.84</td>
<td>8.82 ± 0.20</td>
<td>( P &lt; 0.0001^{****} )</td>
</tr>
<tr>
<td>p_high(ADC)</td>
<td>0.249 ± 0.059</td>
<td>0.073 ± 0.008</td>
<td>( P &lt; 0.0001^{****} )</td>
</tr>
</tbody>
</table>

| PDFF\_mean   | 42.3 ± 2.6    | 43.9 ± 2.0    | \( P = 0.657 \) | 0.505 ± 0.088 |
| PDFF\_median | 42.6 ± 2.5    | 44.3 ± 2.1    | \( P = 0.646 \) | 0.514 ± 0.089 |
| PDFF\_25     | 32.8 ± 2.7    | 37.1 ± 1.9    | \( P = 0.216 \) | 0.588 ± 0.087 |
| PDFF\_10     | 23.5 ± 2.9    | 30.3 ± 1.9    | \( P = 0.060 \) | 0.658 ± 0.086 |
| p\_low(PDFF) | 0.211 ± 0.047 | 0.168 ± 0.031 | \( P = 0.452 \) | 0.586 ± 0.085 |

Table 8 - Comparison of inflammatory parameters between inflamed and non-inflamed patients. Averages are displayed as mean ± standard error of mean; AUC values are shown as value ± standard error. The highest ROC AUC value for the evaluation of inflammation is shown in bold.
### 6.3.2 Structural Parameters

30/53 patients met the criteria for the presence of fat metaplasia, and 23/53 patients did not. Patients with fat metaplasia were significantly older than those without fat metaplasia (mean ages ± SEM were 19.6 ± 0.55 and 17.6 ± 0.54 respectively, \( P = 0.046 \)); there was no significant difference in age between the two groups (\( P = 0.56 \)).

Comparisons of quantitative parameters between patients with and without fat metaplasia are shown in Figure 6-6, and results of the corresponding t-tests and ROC analyses are shown in Table 9. PDFF-based parameters were all significantly increased in patients with fat metaplasia compared to controls. The separation between patients with and without fat metaplasia was improved for parameters which specifically sampled the upper end of the PDFF distribution (i.e. PDFF\(_{75}\), PDFF\(_{90}\) and \(p_{\text{high}}(\text{PDFF})\)). The best performing parameter was \(p_{\text{high}}(\text{PDFF})\), which produced an AUC value of 0.814.

There was no difference in mean or median ADC measurements between patients with and without fat metaplasia.
Table 9 - Comparison of structural parameters between patients with and without fat metaplasia. Averages are displayed as mean ± standard error of mean; AUC values are shown as value ± standard error. The highest ROC AUC value for the evaluation of fat metaplasia is shown in bold.

### 6.3.3 Relationship between qMRI parameters and Visual Scores

The relationship between visual scores of inflammation/fat metaplasia and qMRI parameters is shown in Figure 6-7; results of the linear regression analysis are displayed on the Figure. There were significant positive relationships between ADC\textsubscript{90} and the visual inflammation score, and between \(p_{\text{high}(\text{PDFF})}\) and the fat metaplasia score. For ADC\textsubscript{90}, the estimated y-intercept (corresponding to a visual inflammation score of 0) was 889mm\(^2\)/s, which was significantly greater than 0 (\(P < \))
0.0001). For $p_{\text{high(PDFF)}}$, there was no significant difference between the estimated intercept value and 0 ($P = 0.276$).

### 6.3.4 Relationship between MRI and Symptoms

Scatterplots showing the relationship between BASDAI scores and visual and quantitative scores of inflammation and fat metaplasia are shown in Figure 6-8.

There was no significant relationship between visual scores of inflammation and any clinical score ($P = 0.59, 0.94, 0.064$ and $0.085$ for BASDAI, BASFI, CRP and ESR respectively), or between ADC$_{90}$ parameters and clinical scores (for ADC$_{90}$ $P = 0.41, 0.18, 0.27$ and $0.67$ for the respective clinical scores).

Interestingly, there was a significant negative relationship between fat metaplasia visual scores and between clinical symptoms ($P = 0.009$ and $0.03$ for BASDAI and BASFI), and a similar relationship was observed for the corresponding qMRI parameter $p_{\text{high(PDFF)}}$ ($P = 0.03$ and $0.05$ for BASDAI and BASFI). There was no significant correlation between either visual or quantitative fat metaplasia scores and CRP or ESR (all $P > 0.05$).

![Figure 6-8 - Relationship between BASDAI symptom scores and imaging measures of inflammation and structural damage (fat metaplasia).](image)
6.3.5 Interobserver Agreement

Statistics for interobserver agreement are shown Table 10. ICC measurements indicated excellent agreement for all assessed qMRI parameters. For ADC-based parameters, agreement was slightly better for ADCmedian than for ADC90, both in terms of ICC measurement and 95% limits of agreement. For PDFF-based parameters, agreement was better for percentile measurements (PDFFmedian and PDFF90) than for $p_{\text{high(PDFF)}}$.

ICC measurements also indicated excellent agreement for visual inflammation scores, although the 95% limits of agreement (0.6 ± 6.4) were relatively wide in comparison to the ASAS definition of active inflammation (which regards a score of ≥ 2 as diagnostic for active inflammation). Interobserver agreement was substantially poorer for fat metaplasia scores, with lower ICC values and wider limits of agreement.

<table>
<thead>
<tr>
<th></th>
<th>ICC</th>
<th>Bland-Altman 95% limits of agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADCmedian</td>
<td>0.943</td>
<td>8.4 ± 111 mm²/s</td>
</tr>
<tr>
<td>ADC90</td>
<td>0.918</td>
<td>-10.1 ± 188 mm²/s</td>
</tr>
<tr>
<td>PDFFmedian</td>
<td>0.974</td>
<td>-0.97 ± 5.08 %</td>
</tr>
<tr>
<td>PDFF90</td>
<td>0.986</td>
<td>-0.31 ± 4.70 %</td>
</tr>
<tr>
<td>$p_{\text{high(PDFF)}}$</td>
<td>0.958</td>
<td>-0.67 ± 8.08 %</td>
</tr>
<tr>
<td>Visual inflammation score</td>
<td>0.944</td>
<td>0.6 ± 6.4</td>
</tr>
<tr>
<td>Visual fat metaplasia score</td>
<td>0.534</td>
<td>-6.8 ± 18.3</td>
</tr>
</tbody>
</table>

Table 10 - Interobserver variability statistics.

6.4 Discussion

Many of the previous studies using MRI to quantify inflammation in sacroiliitis have been limited to using manually-placed ROIs on individual inflammatory lesions, and have not attempted to derive a measure of inflammation for the whole joint. This may be partly because inflammatory changes in the subchondral bone marrow can be markedly heterogenous, meaning that simple averages provide a poor representation
of the burden of active inflammation (patients with mixed oedema and fat metaplasia might actually have normal overall measurements). Here, we propose a simple, partially-automated tool - the BEACH tool – which enables an overall measurement of sacroiliac joint inflammation in a reproducible fashion and can separately quantify the active and chronic changes in the marrow. Using the BEACH tool, the observer is only required to identify the line of the sacroiliac joint and the ROIs themselves are generated automatically, removing much of the subjectivity which has afflicted previous studies. The described tool can be regarded as a prototype, and we anticipate that subsequent versions will incorporate automated methods for segmenting subchondral bone, for quantifying disease and for distinguishing between active and chronic inflammation.

In this study, we adopted a histographic approach to measuring inflammation, whereby specific measurements at the extremes of measurement distributions (e.g. 90th centile of ADC measurements for inflammation) are used to separate the active and chronic or ‘structural’ elements of the inflammatory process. We showed that using the 90th centile of ADC measurements yielded a more accurate separation of inflamed and non-inflamed patients compared to simple averages, such as the mean and median. This histographic approach also produced improvements in the separation of patients with and patients without fat metaplasia, and the proportion of high-PDFF pixels - \(p_{\text{high}(\text{PDFF})}\) – produced better performance than either mean or median PDFF values in differentiated inflamed and non-inflamed SIJs. These results suggest that histographic analysis could be a valuable addition to methods for quantifying inflammation in spondyloarthritis and in other inflammatory disease affecting the skeleton. Furthermore, the principles underlying this method are general and could also be applied to other bone marrow diseases – for example haematological disorders such as multiple myeloma – where changes in marrow composition can also be bi-directional [FF measurements decrease as the marrow becomes infiltrated and increase with treatment, potentially to levels which exceed those of normal marrow (305)].

Interestingly, we found that ADC measurements produced superior performance to PDFF measurements in terms of separating patients with and without inflammation. This result suggests that increases in diffusivity are an important part of the inflammatory process in the bone marrow, and may be more important than changes in water content per se. However, previous studies have shown substantial differences
in PDFF between normal and inflamed marrow (see Chapter 5) (306), and it may be that the lack of difference between normal and inflamed SIJs in this study is due to variations in the composition of bone marrow. It has previously been demonstrated that ADC measurements vary with skeletal maturity (263) – probably because water diffusivity is higher in unossified bone compared to fully ossified bone – and it seems likely that PDFF measurements would vary in a similar fashion. This could be investigated further by comparing the composition of the inflamed subchondral bone with normal bone marrow (for example in interforaminal sacral bone, which is often taken as a reference) – it may be that the difference in PDFF values would be more useful than the absolute PDFF values which were used in this study.

Our results suggest that PDFF measurements are particularly useful as a marker of structural damage. The parameter $p_{\text{high(PDFF)}}$ enabled accurate separation of patients with and without fat metaplasia, and also provided faithful measurements of the overall quantity of fat metaplasia in the SIJs. Importantly, although bone marrow oedema is regarded as the key diagnostic finding in spondyloarthritis, the presence of fat metaplasia can help to improve diagnostic confidence and is also an important prognostic factor, since patients with fat metaplasia are more likely to fuse their sacroiliac joints (56,108,115).

We did not find a strong relationship between inflammation on MRI and symptoms in this study, although this is perhaps unsurprising given the complex and multidimensional nature of pain in spondyloarthritis, and in young spondyloarthritis patients in particular (133). Interestingly, we observed a negative relationship between the severity of fat metaplasia and symptom scores (symptom scores were higher in patients without fat metaplasia), possibly because chronic pain is more common in the control population than in patients with spondyloarthritis. This relationship might also reflect the fact that fat metaplasia - which is thought to be a post-inflammatory phenomenon (55,115) – is more common in patients already on treatment, who therefore have well-controlled symptoms.

A limitation of this study is that, in young patients, the polygonal ROIs propagated on subchondral bone are likely to include the unossified intersegmental sacral apophyses, which return predominantly water signal with very little fat. This may bias the qMRI measurements in those patients and could have weakened the separation of qMRI parameters between inflamed and non-inflamed patients, as
discussed above. In the future, the BEACH tool could be extended to include a segmentation algorithm which specifically isolates ossified bone, which could potentially improve performance substantially. Similarly, variations in the proportion of red and yellow in the ossified bone may vary between individuals, and the use of variable thresholds depending on the composition of the normal ‘background’ marrow might help to improve the sensitivity and specificity of the technique for detecting inflammation. As discussed in Chapter 3, ADC measurements can also suffer from poor reproducibility across sites, partly due to the difficulty of achieving high-quality fat suppression. Finally, it should be highlighted that the diagnosis of spondyloarthritis is a complex process depending on clinical assessment and blood tests in addition to MRI results, and the patients who were deemed to have inflammation on their scans on this study do not necessarily meet the overall clinical criteria for diagnosis. Conversely, some patients who did not meet the MRI criteria for inflammation may have subclinical inflammation, which could have influenced their qMRI scores. To examine this further, it would be useful to compare the qMRI data from the subjects in this study with those from an age-matched or skeletal maturity-matched healthy control group.

6.5 Conclusion

The BEACH tool enables simple, partially-automated measurements of both active inflammation and structural damage in the subchondral bone marrow of the sacroiliac joints, and eliminates much of the subjectivity which has afflicted previous studies aiming to quantify inflammation in the sacroiliac joints. This tool incorporates histographic analysis of the subchondral bone, which provides superior performance compared to simple averaging in terms of separating inflamed joints from non-inflamed joints and separating joints with fat metaplasia compared to those without. The described method can be regarded as a prototype, and we anticipate that future versions of this tool will involve fully automated methods for segmenting bone and for quantifying inflammation and structural damage.
6.6 Supplementary information

6.6.1 Generation of polygonal ROIs

The observer is prompted to define the line of the sacroiliac joint using a single series of connected straight lines - an open polygon – as shown in Figure 6-9. This open polygon includes ‘anchor lines’ (shown in blue), which define the angle made by the joint with the cortical surface, at both the top and bottom of the joint (Figure 6-9), and the line of the joint itself. The first anchor line is placed perpendicular to the surface of the bone at the upper end of the sacroiliac joint, in order to define the angle $\alpha$ between the bone surface (labelled $s$) and the joint itself. Next, the observer places a series of joint lines (also shown in blue, labelled $j_1 \ldots j_n$), which trace the course of the synovial part of the sacroiliac joint. Finally, the observer places a second anchor line defining the angle $\beta$ formed between the joint and the cortical surface (in the case of axial diffusion-weighted images, the ligamentous part of the joint is not including in the scoring so the angle $\beta$ is formed between the synovial and ligamentous parts of the joint, rather than at the cortical surface). A series of lines (shown in red on Figure 6-9a) with width $w_i$ are then automatically generated and placed such that each line is perpendicular to the preceding joint line. In this study, we used $w_i = 14$ pixels (equal to 21.8mm for PDFF maps and 23.4mm for PDFF).

![Figure 6-9 - Geometry of ROIs propagated on subchondral bone.](image-url)
ADC maps), chosen empirically. The coordinates of these lines are used to form a polygonal ROI including both the subchondral bone and joint space (on Figure 6-9b). This procedure is repeated to generate a further set of shorter perpendicular lines, with width \( w_2 \), which define the width of the joint space itself (not shown). Here, we used \( w_2 = 6 \) pixels (equal to 9.36mm for PDFF maps and 10.02mm for ADC maps). Finally, the coordinates of these two sets of lines are used to generate two separate polygonal areas \( A_1 \) and \( A_2 \) adjacent to the joint, which comprise the total region-of-interest \( A \) for a single SIJ (on Figure 6-9c).

6.6.2 Derivation of PDFF and ADC Thresholds

To derive PDFF thresholds for inflammation and fat metaplasia, the data from the ROIs placed on areas of oedema, fat metaplasia in normal marrow in Chapter 5 (306) underwent ROC analysis. The optimal PDFF cutoff for differentiating oedema from normal marrow or fat metaplasia was 28.2%, giving a sensitivity of 89% and a specificity of 92% (AUC=0.95). The optimal PDFF cutoff for differentiating fat metaplasia from normal marrow or oedema was 69.7%, giving a sensitivity of 100% and a specificity of 100% (AUC=1). These thresholds were incorporated into the software for the BEACH tool to derive the qMRI parameters described in Section 6.2.4. To derive an ADC threshold for separating oedema from normal marrow, ROIs were placed on ADC maps on an independent cohort of patients with sacroilitis (\( n=30 \), consisting of 20 patients with spondyloarthritis, meeting the ASAS criteria for active inflammation, and 10 controls), which was identified using a picture archiving and communications (PACS) database search. In patients with sacroilitis, freehand ROIs were placed on up to six regions of bone marrow oedema, which were identified on \( T_2 \)-weighted short tau inversion recovery (STIR) images as areas of hyperintense signal (119). In controls, freehand ROIs were placed on six regions of normal subchondral bone marrow, taking care to avoid the joint and adjacent soft tissue. In patients with evidence of chronic sacroilitis (i.e. those with fat metaplasia, joint fusion or erosions) but no active inflammation, no ROIs were placed. The ROC analysis revealed an optimal ADC cutoff of 957mm\(^2\)/s, which produced a sensitivity of 96.6% and specificity of 100% (AUC = 0.99).
Quantitative Susceptibility Mapping for Measurement of Bone Mineral Density

This work was jointly performed by this author (TB) and Anita Karsa (AK). TB conceived of the project, designed and constructed the phantoms, acquired the images and placed the ROIs. AK processed the images to generate susceptibility maps, derived the ROI measurements and performed the regression analyses. Both TB and AK designed the analysis plan and drafted the manuscript.

7.1 Introduction

New bone formation is a key feature of spondyloarthritis and causes spinal fusion, which contributes to pain, morbidity and disability. Conversely, spondyloarthritis patients may also suffer from bone loss in the form of osteoporosis (62), which contributes to increased fracture risk. Both disease processes cause alterations in bone mineral density (BMD), but this tissue property is difficult to measure using conventional MRI (119,163). As such, there is a need for a quantitative MRI-based method which can be used to measure BMD, and therefore monitor new bone formation and bone loss in spondyloarthritis.

In Chapter 5, $R_2^*$ was investigated as a quantitative biomarker of trabecular bone mineral density (BMD) as the diamagnetic nature of bony trabeculae is expected to increase the rate of signal decay. We found a correlation between BMD and $R_2^*$ in a fat-water-bone phantom (a mixture of peanut oil, agar solution, and granules of bovine bone matrix), and also significantly reduced $R_2^*$ in areas of fat metaplasia in patients with spondyloarthritis. However, $R_2^*$ measurements were also influenced by variations in fat content, and the relationship between fat fraction (FF) measurements and $R_2^*$ was complex. This complexity probably arises because fat contributes to dephasing both within the voxel (due to ‘spectral broadening’ effects arising from the spectral complexity of fat) and in adjacent voxels, due to dipole patterns arising from susceptibility differences between fat and water (fat has been repeatedly found to be more paramagnetic than water-based tissue (307–309)). Furthermore, $R_2^*$ measurements cannot differentiate between para- and diamagnetic structures (310).
Recently, quantitative susceptibility mapping (QSM) (310–312) has been investigated as an alternative method for quantifying BMD, with promising initial results (313). Dimov et al. showed that susceptibility values were closely correlated with CT measurements in a porcine hoof, and were able to generate susceptibility maps in which cortical bone was homogenous and diamagnetic, as expected from theory (313). However, susceptibility mapping is challenging in the presence of varying fat content, which is a characteristic feature of bone marrow inflammation in spondyloarthritis (306). Like $R_2^*$ measurements, susceptibility estimates can be confounded by variations in fat content, which contribute to dephasing both within the voxel (due to chemical shift) and in adjacent voxels (due to local dipole patterns arising from the higher susceptibility of fat relative to fluid-based tissue).

In this study, we investigated the feasibility of using QSM to measure BMD in inflamed bone marrow. The described method was designed to correct for the effect of chemical shift, and we also attempted to separate the fat contribution to total susceptibility such that ‘fat-corrected’ susceptibility measurements could be calculated. We evaluated the relationship between FF and BMD in dedicated phantoms containing fat, water and trabecular bone, and also evaluated differences in susceptibility between areas of normal marrow, oedema and fat metaplasia in patients with spondyloarthritis.

7.2 Methods

This study received ethical approval from the Queen Square Research Ethics Committee, London, United Kingdom (Research Ethics Committee reference 15/LO/1475). All patients gave written informed consent prior to study entry.

7.2.1 Phantoms

The effect of fat fraction (FF) and bone mineral density (BMD) on susceptibility was investigated using the fat-water-bone phantom described in Chapter 5, which consists of varying concentrations of peanut oil, water and decellularised bovine trabecular bone matrix. A new fat-water phantom was also created to examine the relationship between susceptibility and fat fraction over the full FF range. Eleven 50-mL centrifuge tubes were filled with mixtures of water and lard (rather than peanut oil), using sodium dodecyl sulphate (SDS) as a surfactant, with FF values
varying from 0% to 100%, in 10% increments. The final SDS concentration in each phantom was 28mM.

7.2.2 Patients and Volunteers

This study was performed using the data from the spondyloarthritis patients reported in Chapter 5. As before, patients with suspected spondyloarthritis were treated as controls if the subsequent clinical MRI scan and clinical assessment were normal.

7.2.3 Data Acquisition

To generate the susceptibility maps, we used the raw complex data (consisting of magnitude and phase images) of the fat-water-bone phantom and subjects rather than the mDixon Quant data reported in Chapter 5. These complex images were acquired using a 3D spoiled gradient-echo pulse sequence with bipolar readout gradients, with integrated posterior and anterior surface coils. Images of the fat-water-bone phantom (and also the new fat-water phantom) were acquired coronally, with the following parameters: matrix size = 320×320×53, resolution = 0.94×0.94×1.5 mm³, TE₁ = 1.233 ms, ΔTE = 1.951 ms, six echoes, TR = 23 ms, flip angle = 3°, bandwith 1159Hz/pixel. Images of patients consisted of tilted coronal slices (parallel to the long axis of the sacrum), a matrix size = 320×320×40, resolution = 1.56×1.56×2 mm³, TE₁ = 1.17 ms, ΔTE = 1.6 ms, 6 echoes, TR = 25 ms, flip angle = 3°, bandwith 1894Hz/pixel. As described in Chapter 5, the mDixon Quant acquisition provided proton density fat fraction (PDFF) and R₂* maps (with the same matrix size and FOV as the raw complex data), and subjects also underwent a standard clinical MRI scan on a 1.5T system (Avanto, Siemens, DE) with angled coronal (tilted at the same angle as the gradient-echo images) T₁- and T₂-weighted STIR sequences. These images were used for manual segmentation of normal bone marrow, bone marrow oedema, and fat metaplasia.

7.2.4 Derivation of Quantitative Susceptibility Maps from Phase Images

The basic goal of susceptibility mapping is to derive the susceptibility distribution from the field map. These are related by
\[ C \chi = \psi \]  \[ \text{[39]} \]

where \( C \) is the matrix representation of convolution with the dipole response, \( \psi \) is the field map, and \( \chi \) is the susceptibility distribution. The necessary deconvolution operation can be achieved by k-space division (310) (a division in k-space is equivalent to deconvolution in image space); however, the problem is ill-posed, meaning that a number of different solutions in \( \psi \) are possible. Furthermore, the acquired phase data are affected by variations in the Bo field, chemical shift effects and phase wrapping. Therefore, a series of steps are typically used to mitigate these effects and to improve the accuracy of the susceptibility estimates; the specific methods used in this work are outlined here.

1. The three-point Dixon method described by Berglund (198) was used to separate fat and water and therefore estimate the Bo map independent of chemical shift effects, as described in Section 2.3.6.3.

2. Phase unwrapping was performed using a method described by Schweser et al. (314), which is based on an earlier method using an earlier method described by Schofield and Zhu (315). In this method, the Laplacian of the unwrapped phase is derived using

\[ \Delta \phi_{\text{unwr}}(r) = L_{\text{wr}} \phi_{\text{wr}}(r) \]  \[ \text{[40]} \]

where \( \phi_{\text{unwr}} \) and \( \phi_{\text{wr}} \) are the unwrapped and wrapped phase images respectively, \( r \) is a spatial coordinate vector, \( \Delta \) denotes the Laplace operator and \( L_{\text{wr}} \) is a nonlinear operator applying trigonometric functions and Laplacians to the wrapped phase image (314). The unwrapped phase image can then be obtained by applying the inverse Laplace operator (denoted \( L^{-1} \)):

\[ \phi_{\text{unwr}} = L^{-1} L_{\text{wr}} \phi_{\text{wr}} \]  \[ \text{[41]} \]

*The Laplacian is a second order differential operator which, in Cartesian coordinates, is equivalent to the sum of the second order partial derivatives in each of the coordinates.*
3. Having generated an unwrapped phase image, a method known as projection onto dipole fields (PDF) was used to remove variations in the B0 field which are not attributable to variations in susceptibility (this step is commonly known as background field removal) \(^{(316)}\). This method considers the relationship between the local field \(f_L\), defined as the magnetic field generated by the susceptibility distribution \(\chi_L\) inside an ROI, and the background field \(f_B\), defined as the magnetic field generated by the susceptibility distribution \(\chi_B\) outside this ROI. These are related by

\[
    f = f_L + f_B = d \otimes [\chi_L + \chi_B]. \tag{42}
\]

where \(d\) is the unit dipole field. The PDF method makes use of the fact that the inner product of the field of a background dipole outside the ROI and the field of a local dipole inside the ROI is almost zero in the ROI, enabling decomposition of \(f_L\) and \(f_B\); further detail is given in \((316)\). The PDF method was shown to offer improvements in background field removal compared to high pass filtering \((316)\), and is now commonly used as part of QSM processing pipelines.

4. Having removed background fields, the contrast in the unwrapped phase image should now be almost entirely due to local variations in tissue susceptibility. Here, the susceptibility maps were derived using a method described by Kressler et al., which incorporates Tikhonov regularization to improve the conditioning of the inverse problem described by Eq. \((39)\) \((317)\). The problem can then be formulated as

\[
    \min_\chi \|W(C\chi - \psi)\|^2_2 + \alpha^2 \|L\chi\|^2_2 \tag{43}
\]

\(\alpha\) is a tunable regularization parameter, \(L\) is the Tikhonov regularization matrix (this may also be denoted \(\Gamma\)), and \(W\) is a weighting matrix designed to reduce noise propagation. The regularization matrix \(L\) gives preference with solutions to smaller \(L_2\) norms (thus the method is also known as \(L_2\) regularisation). This approach improves the conditioning of the problem and enables a direct numerical solution. It can be viewed as a refinement of the
method described previously by Shmueli et al., where the susceptibility maps were truncated at a fixed threshold to remove erroneously high values due to noise amplification introduced by the k-space division (310).

Note that the three-point Dixon method (step 1) requires only three equally spaced echoes. We used the first, third, and fifth echoes of both the phantom and subject images, as these consistently provided images with the fewest fat-water swapping artifacts by visual inspection. Additionally, all images were zero-padded to a matrix size of $512 \times 512 \times 128$ before steps 2 and 4. The tilt of the coronal slices was accounted for by defining the dipole kernel to be parallel to the real direction of the main magnetic field in steps 3 and 4. The Tikhonov regularisation parameter was set to $\alpha = 0.05$ in step 4.

The background field removal (step 3) requires a binary tissue mask. Initial masks were obtained in each case by thresholding the inverse noise map calculated from the multi-echo magnitude images (317,318) to exclude high-noise voxels that could introduce streaking into the susceptibility maps. In the phantoms, artifact-inducing structures were manually segmented in the first-echo magnitude images using ITK-SNAP (319,320) and also excluded from the mask. In the patient and volunteer images, bony voxels were not excluded from the tissue mask despite their low signal. These bony voxels were identified in all subjects using the following scheme: 1. Bones were manually segmented in the first-echo magnitude image of one of the healthy volunteers (subject 1) in ITK-SNAP (319,320). 2. All scanner-provided water images were thresholded so that values in regions with low water signal were set to zero. 3. The thresholded water image of subject 1 was non-rigidly registered to all other thresholded water images using the NiftyReg software (321) with the weight of the bending energy term increased to 0.01 and a final grid size of 7 voxels. 4. Bones were segmented in the rest of the images by applying the resulting transformations to the manually segmented bone region of subject 1. This process provided suitable segmentations in all subjects. We used the thresholded water images for step 3, because the shape and size of subcutaneous fat largely varied across subjects while the water images generally looked similar, therefore provided more accurate registrations around bony structures. Additionally, the edges of the patient and volunteer tissue masks were eroded by 5 voxels in each slice to further improve the quality of the susceptibility maps.
In theory, QSM calculates the bulk susceptibility in each voxel which is expected to have a linear relationship with both PDFF and bone mineral density (BMD). Therefore, we propose a procedure to estimate BMD-induced susceptibility maps:

1. Susceptibility and PDFF maps are calculated.
2. Linear regression is performed between susceptibility and PDFF measured in voxels without bony trabeculae.
3. The regression parameters and the PDFF map are used to estimate the contribution of fat to susceptibility in every voxel.
4. The contribution of fat is subtracted from the total susceptibility map resulting in a susceptibility map that is supposed to be proportional to BMD assuming that no other paramagnetic or diamagnetic components are present.

We performed this procedure in all volunteer and patient susceptibility maps using the scanner-provided PDFF maps. Step 2 was carried out in a rectangular region, including both water-based tissue (gluteal muscle) and subcutaneous fat, manually selected in the middle slice in each subject. A similar procedure was performed for the scanner-provided R₂* maps. Since it is very difficult to accurately model the effects of PDFF and BMD on the measured R₂*, we adopted a quadratic instead of a linear fit in Step 2 that seemed to provide acceptable fits in all subjects by visual inspection.

### 7.2.5 Statistical Analysis

For both phantoms, circular regions of interest (ROIs) were manually drawn (by A.K.) on the first-echo magnitude images in eight consecutive slices near the middle of the acquired volumes using ITK-SNAP (319,320). Mean susceptibilities were calculated in all ROIs. A 2D linear function was fitted to the measured susceptibility values as a function of known FF and BMD values in the fat-water-bone phantom. Linear regression was performed between measured susceptibilities and known FF values in the fat-water phantom.

For the patients and healthy controls, areas of normal bone marrow, bone marrow oedema, and fat metaplasia were manually segmented on the T₂-weighted STIR and T₁-weighted images and then transferred to PDFF and susceptibility maps by a radiology registrar (TB) as described in Chapter 5. ROIs which were very close to fat-water chemical shift artifacts in the susceptibility maps were excluded from the
analysis (this included all ROIs from two subjects, and two additional ROIs from a third subject). Mean susceptibilities and $R_2^*$ values were calculated in the rest of the segmented ROIs both before and after removing the contributions of fat. Susceptibility values were referenced to the mean susceptibility within the tissue mask for each subject. Multi-level mixed-effects linear regression was used (in Matlab R2015a) to determine whether there were significant differences in susceptibilities and $R_2^*$ values measured in normal bone marrow, oedema, and fat metaplasia. This test accounts for repeated observations in individual patients.

7.3 Results

7.3.1 Phantoms

Images from the fat-water-bone phantom are shown in Figure 7-1a,b. Susceptibility measurements were positively related to FF values and negatively related to BMD (Figure 7-1c), with the 2D linear model providing an accurate description of the acquired data (in Figure 7-1c-g). Similarly, in the lard-water phantom covering the full range of FF values, there was an approximately linear relationship between FF and susceptibility (Figure 7-2).

7.3.2 Patients and volunteers

Susceptibility values were significantly increased in areas of fat metaplasia compared to normal marrow (Figure 7-3). $R_2^*$ measurements were also significantly reduced in areas of fat metaplasia compared to normal marrow (Figure 7-4), in accordance with the previous results (see Chapter 5) (7).

PDFF and susceptibility values within a single, rectangular ROI (overlaid on the susceptibility map) incorporating both muscle and subcutaneous fat, in addition to the results of the linear regression analysis, are shown for a single subject in Figure 7-5. Similarly, Figure 7-6 shows $R_2^*$ and PDFF values within the manually selected ROI (overlaid on the $R_2^*$ map), and the results of the nonlinear regression analysis assuming a quadratic relationship. Model parameters from the linear and quadratic fits between PDFF and susceptibility, and PDFF and $R_2^*$ respectively are shown in Figure 7-7. The coefficients of the quadratic fit (Figure 7-7b) largely varied across subjects. While the intercept of the linear fit also showed large variations across subjects, the slope was somewhat consistent for regressions of high adjusted $R^2$ measures (Figure 7-7a, blue circle). Susceptibility values and $R_2^*$ measurements
Chapter 7 - Quantitative Susceptibility Mapping for Measurement of Bone Mineral Density

After removing the fat contribution are shown in Figure 7-8 and Figure 7-9 respectively. There was no significant difference in either susceptibility or $R^*_2$ measurements between fat metaplasia and normal marrow after performing the adjustment for fat content.

![Diagram](image)

Figure 7-1 - Results from the fat-water-bone phantom. First-echo magnitude image and susceptibility map are shown in (a) and (b) respectively. The manually drawn circular ROIs are highlighted in red (a). Results of the 2D linear fit between bone mineral density (BMD) and fat fraction (FF) values and susceptibility are shown in (c-g). In (c), the transparent surface corresponds to measured values, while the opaque plane is the fitting 2D linear function.
Figure 7-2 - Results from the fat-water phantom. First-echo magnitude image and susceptibility map are shown in (a) and (b) respectively. The linear fit between fat fraction (FF) and susceptibility values is shown in (c-d).
Figure 7-3 - Susceptibility maps in patients. Measured mean susceptibilities in areas of normal marrow, oedema, and fat metaplasia are shown in (a). p-values were calculated for each pair and the asterisks indicate statistical significance. Susceptibility maps and magnitude images in example subjects are shown in (b).

Figure 7-4 - R₂* maps in patients. Mean R₂* measurements in areas of normal marrow, oedema, and fat metaplasia are shown in (a). p-values were calculated for each pair and the asterisks indicate statistical significance. R₂* maps and magnitude images in example subjects are shown in (b).
Chapter 7 - Quantitative Susceptibility Mapping for Measurement of Bone Mineral Density

Figure 7-5 - Linear regression between PDFF and susceptibility values in a single, representative subject. A rectangular ROI including fat and muscle was placed on the middle slice of the susceptibility map. Each point in the scatter plot corresponds to one voxel within this rectangular region. The blue arrow and dotted circle indicate subcutaneous fat and the corresponding points in the scatter plot. The green arrow and dotted circle indicate muscle and the corresponding points in the scatter plot. Results from the regression analysis for each subject are shown in Figure 7-7.

Figure 7-6 - Nonlinear regression between PDFF and $R_2^*$ values using a quadratic model. The $R_2^*$ map and the regression of a representative subject are shown. Each point in the scatter plot corresponds to one voxel within this rectangular region. The blue arrow and dotted circle indicate subcutaneous fat and the corresponding points in the scatter plot. The green arrow and dotted circle indicate muscle and the corresponding points in the scatter plot. Results from the regression analysis for each subject are shown in Figure 7-7.
Figure 7.7 - Coefficients of the linear and quadratic fits between PDFF and susceptibility (a) and \( R_t^* \) (b) for each of the subjects included in the study. The fitting function is displayed in the top right corner of both subplots. In all five scatter plots, each point corresponds to one subject. Coefficients are shown as a function of the adjusted \( R^2 \) of each fit. The error bars indicate the 95\% confidence interval of each coefficient. The slope of the linear regression (a) seems to be consistent in instances where the adjusted \( R^2 \) was high (>0.5). All other parameters had large variations across subjects.
After removing the fat contributions to susceptibility:

a) Measured susceptibilities in areas of normal bone marrow, oedema and fat metaplasia

b) Susceptibility maps and Magnitude images

Normal bone marrow (●)
Oedema (●)
Fat metaplasia (★)

Figure 7-8 - Susceptibility measurements in patients after removing the fat contribution.

After removing the fat contributions to R2*:  

a) Measured R2* in areas of normal bone marrow, oedema and fat metaplasia

b) R2* maps and Magnitude images

Normal bone marrow (●)
Oedema (●)
Fat metaplasia (★)

Figure 7-9 - R2* measurements in patients after removing the fat contribution.
Chapter 7 - Quantitative Susceptibility Mapping for Measurement of Bone Mineral Density

7.4 Discussion

New bone formation and bone destruction contribute to spinal ankylosis and osteoporosis respectively in spondyloarthritis, and are key contributors to morbidity. However, these processes cannot be easily quantified using conventional spin echo sequences. In this study, we sought to characterize the relationship between bone mineral density (BMD), fat fraction (FF), and susceptibility measurements in inflamed bone marrow, using both phantom and in vivo studies.

In the fat-water-bone phantom, we observed linear relationships between FF and susceptibility and between BMD and susceptibility. The linear relationship between FF and susceptibility was also observed in a separate, lard-based fat-water phantom covering the full range of FF measurements. In accordance with previous studies (307–309,313), our data indicated positive (paramagnetic) susceptibility values for fat, and negative (diamagnetic) susceptibility values for bone. These results confirm the feasibility of measuring BMD in the bone marrow using the proposed method, and suggest that the contribution of fat to the total susceptibility measurement can be removed using a simple linear relationship.

Importantly, the results of our phantom study were used to inform the analysis of the in vivo results, where we were able to estimate the contribution of the fat to the total susceptibility measurement. Using per-patient linear regression analysis in voxels of subcutaneous fat and muscle, we could remove the fat contribution to susceptibility measurements in areas of fat metaplasia, oedema and normal marrow, and thereby interrogate the source of susceptibility differences between these regions. Strikingly, we found that susceptibility measurements were significantly increased in areas of fat metaplasia compared to normal marrow, but that this difference was abolished after removal of the fat contribution. This suggests that the contribution of fat content to overall susceptibility is likely to be substantial, and highlights the importance of accounting for the contribution of fat when performing QSM in bone marrow.

Similarly, although there was a significant reduction in $R_2^*$ in areas of fat metaplasia compared to normal marrow, no significant difference was observed in fat-corrected $R_2^*$ measurements. This result implies that the previously-reported reduction in $R_2^*$ (Chapter 5) in areas of fat metaplasia may actually be a secondary effect of varying
fat content: as fat fraction increases from around 50% (normal bone) to about 90% (fat metaplasia), the susceptibility distribution inside the voxel becomes more homogenous, and the relaxation rate $R_2^*$ reduces accordingly. This suggestion is in keeping with previous results in muscle, which suggest that $R_2^*$ measurements are highest at intermediate FF values, and are at their lowest at the extremes of the PDFF range (i.e. close to 0% and close to 100%) (284).

Here we used the individual regression parameters for each subject for removing the contributions of fat from susceptibility and $R_2^*$ maps. In case of the $R_2^*$ maps, the quadratic model is a heuristic approximation of the observed shape of the PDFF-$R_2^*$ relationship without any physical reasoning behind it. Therefore, the coefficients largely vary across subjects, and the correction is expected to be more accurate if individual fitting parameters are used. Moreover, since susceptibility maps provided by QSM are only defined up to an offset, the intercept of the linear model is also expected to have large inter-subject variations (although in theory the slope of the linear regression should be consistent). Therefore, using individual linear regression parameters is expected to provide more accurate ‘fat-correction’ and enable a more unbiased comparison between the susceptibility and $R_2^*$ measurements. However, since the estimated slope seems to be somewhat consistent across subjects for regressions with high adjusted $R^2$, it could be interesting to explore if a single slope value can be used to robustly remove fat contributions from susceptibility maps.

Overall, the results of our study highlight the importance of removing the fat contribution from susceptibility and $R_2^*$ estimates. Without this correction, changes in susceptibility/$R_2^*$ might be incorrectly attributed to changes in BMD (or other factors). The fact that we did not find a significant difference in fat-corrected susceptibility argues against a significant change in BMD in patients with spondyloarthritis; however, it is also possible that we have simply failed to detect this change due to technical limitations arising from the susceptibility mapping pipeline. Further improvement of the acquisition, the susceptibility mapping pipeline, and in the sample sizes used for this study may help to address this issue more definitively.

Noise amplification is an inherent problem with QSM and arises largely because of the inverse and ill-posed nature of the problem. Accordingly, the susceptibility measurements in this study had a relatively large variance, even in normal bone
marrow. Some groups have previously investigated methods for reducing noise effects in the contexts of brain imaging (322), but there has been little or no comparable research on QSM methods in the body. Further research is needed to compare candidate QSM methods in the context of bone imaging, and to evaluate the relative contribution of the various processing steps to the total variance of QSM measurements.

Some specific aspects of the processing pipeline deserve particular attention. The fat-water decomposition step in the QSM pipeline, aiming to eliminate chemical shift effects, suffered from fat-water swaps in some subjects, which may have contributed to inaccuracies in the measured susceptibilities. This may be due to phase errors arising from the multi-echo readout used to acquire the data or from errors related to the Berglund method used here. Empirically, we found that this method was the most robust of the available options in the ISMRM fat-water toolbox (27), although even this did not perform perfectly in all cases. Better results might be achieved by using alternative algorithms for fat-water decomposition. One option is to use manufacturers’ own algorithms for fat-water decomposition (and to generate field maps), but this comes at the cost of reduced flexibility and makes it more difficult to translate the approach to other systems. Another possibility is using in-phase echo timing to remove most of the chemical shift phase contributions, while also acquiring opposed-phase (or partially opposed-phase) images to calculate PDFF maps.

One of the most crucial features of susceptibility mapping is the generation of the tissue mask. Noisy voxels are prone to introduce far-reaching streaking artifacts and errors into the susceptibility maps. In images of the sacroiliac joint, it is very important to properly exclude areas of bowel as the phase measured in these voxels is often corrupted by motion artifacts and suffers from low signal due to the presence of air. Thresholding the inverse noise map is a wide-spread method for generating a suitable tissue mask. However, bony voxels are also expected to have low signal. The process described here aiming to keep bony voxels while excluding the bowels was simple and provided reasonable tissue masks in most cases, but susceptibility accuracy could potentially be improved using more accurate bone segmentation.

Furthermore, susceptibility maps are only defined up to an offset. Therefore, the measured mean susceptibilities are generally referenced to a value that is not
expected to have large inter-subject variations. Here we used the mean susceptibility within the entire tissue mask as reference. Future studies could explore other potential reference tissues in the pelvic area.

In general, using susceptibility as a marker of BMD has several advantages over $R_2^*$. Most importantly, the linear relationships observed in the fat-water-bone phantom enable fat-correction to be performed very simply, and the fat-corrected susceptibility measurements to be interpreted unambiguously. By comparison, there is a complex relationship between PDFF and $R_2^*$ measurements, which means that changes in $R_2^*$ in tissue are ambiguous. It might be possible to model this relationship using prior knowledge of fat and water susceptibility and the arrangement of fat and water in the tissue, but this is not trivial and introduces further sources of complexity. Nonetheless, there is clearly scope for further optimisation of the susceptibility mapping method described here.

### 7.5 Conclusion

Quantitative susceptibility measurements are linearly related to both bone mineral density (BMD) and fat fraction (FF), and failure to remove the fat contribution to susceptibility measurements can potentially lead to errors in BMD quantification. We propose a method for removing this contribution. Comparison of data both with and without this correction suggest that the increase in susceptibility in areas of fat metaplasia is at least partly due to changing fat content.
8 Discussion and Perspectives

8.1 Overview

In this thesis, two main qMRI techniques were investigated as potential methods for quantifying skeletal inflammation.

The first of these, diffusion-weighted imaging, has been investigated previously and offers an intuitive approach to measuring inflammation. By definition, inflammatory exudates cause an expansion of the extracellular space and thus an overall increase in the freedom of water diffusion, which can be measured using DWI. In this work, we have shown that ADC measurements are responsive to changes in inflammation, and can distinguish between patients with and without inflammation of the SIJs. Importantly, the results of the BEACH study (Chapter 6) suggest that specifically sampling the upper end of the ADC distribution can improve diagnostic performance in terms of separating patients with and without SIJ inflammation. This is a logical result – if 10% of the subchondral bone is inflamed the 90th centile would be expected to increase whereas the median would be expected to change very little – and could potentially improve the way that qMRI is used to quantify inflammation in future studies. In the BEACH study we also found that DWI was superior to CSE-MRI in terms of separating inflamed and non-inflamed SIJs. This implies that the increase in water diffusion is an important aspect of tissue pathology in spondyloarthritis, and may be more important than increasing water content per se.

However, it is important to highlight that DWI does have some limitations, including the confounding effect of fat, problems with image resolution and the presence of artifacts introduced by the EPI readout. These problems are not insurmountable, but more research is needed to improve DWI acquisitions and to standardize ADC measurements across sites. Alternatively, inflammatory exudates might be better characterised using T2 mapping, since ADC and T2 measurements both depend on the mobility of water molecules (this idea is discussed further below, see Section 8.2.1). Either way, it seems likely that a technique specifically evaluating the change in the mobility of water molecules will continue to be an important component of inflammatory imaging protocols.
The second method investigated, CSE-MRI, has not been used previously in the context of skeletal inflammation and represents a different approach. With this technique, we rely on the fact that normal bone marrow has a fat fraction value of around 50%; this value decreases in areas of bone marrow oedema (the fluid can be thought of as 'displacing' the fat) and increases in areas of fat metaplasia (which can be regarded as a form of structural damage). A fundamental difference compared to DWI is that, rather than using fat suppression (i.e. inversion recovery or chemical fat saturation methods), we simply acquire the fat signal and separate it from the water signal using post-processing. Since the fat fraction is a proportion, it is effectively standardized and is therefore relatively immune to differences in coil sensitivity and sequence parameters, which contributes to the excellent reproducibility of this parameter.

Despite the clear reduction in PDFF in focal areas of bone marrow oedema (Chapter 5) we did not find a significant overall reduction in PDFF in patients with SIJ inflammation. This surprising result might indicate that the changes in PDFF occurring due to oedema are actually relatively small (for example, compared to changes in ADC measurements). Alternatively, it may be that the wide variation in normal bone marrow composition between individuals (due to differences in skeletal maturity and in the proportion of red and yellow marrow) obscured the effect of inflammation in this study. To evaluate this further, it would be useful to examine the relationship between the 'background' marrow and the inflamed subchondral bone in patients with spondyloarthritis. It is possible that adjusting for the background composition may help to 'unmask' the inflammatory signal, and therefore improve the performance of PDFF as a marker of inflammation. This idea is discussed further below (Section 8.2.2).

In both the initial proof-of-principle study and the BEACH study (Chapters 5 and 6), PDFF measurements could more accurately distinguish between patients with and without fat metaplasia than those with and without inflammation. This may be because the changes in PDFF are larger in areas of fat metaplasia than oedema, particularly in young patients with water-dominant marrow. These data suggest that PDFF measurements (particularly PDFF\text{90} and p_{high(PDFF)}) might be useful as markers of structural damage in spondyloarthritis. Measurements of structural damage may be important as an outcome measure, and might also have some prognostic significance, for example in prediction of response to biologic therapy.
Finally, several methods for quantifying bone mineral density – and thus new bone formation and destruction – were evaluated in this thesis. Encouragingly, both R2* and susceptibility measurements could detect differences in bone mineral density in fat-water-bone phantoms. However, the results in vivo were less clear cut, and both measurements were influenced by variations in fat content in addition to BMD. Both techniques have their own advantages. R2* mapping is the simpler technique, but cannot differentiate between paramagnetic and diamagnetic structures and has a complex relationship with fat content. Conversely, the post-processing required for QSM is more complex, but susceptibility measurements can differentiate paramagnetic and diamagnetic structures and demonstrate simpler, linear relationships with BMD and fat content (enabling correction for varying fat content to be performed more easily). Overall, it remains unclear whether bone mineral density is altered in areas of fat metaplasia, and more work is needed to investigate methods for quantifying bone formation in spondyloarthritis.

In summary, the work presented in this thesis describes several potential approaches to quantifying inflammation and structural damage, with promising initial results. However, quantitative imaging of inflammation is still in its infancy and there are a number of areas where improvement is needed - both in terms of acquisition and post-processing strategies – before these methods will be clinically useful. The remainder of this Chapter addresses specific outstanding challenges, and discusses potential technical developments which might improve our ability to detect, quantify and characterize skeletal inflammation using MRI. The path to eventual clinical implementation is also discussed.

8.2 Directions for Further Research

The following Section describes potential areas for development. We have already begun preliminary work on some of these areas (particularly 8.2.1 8.2.2, and 8.2.3); and several other potential areas for development are highlighted (see 8.2.4, 8.2.5)

8.2.1 Improved Imaging of Oedema

Although ADC measurements can be used to quantify inflammation, they are dependent on the quality of fat suppression and suffer from problems with reproducibility across sites. Conversely, PDFF measurements can also be used to detect oedema and are more reproducible than ADC measurements, but we found
that PDFF performed less well than ADC in differentiating patients with inflamed SIJs from those without SIJ inflammation. A limitation of using the proton density fat fraction is that the conventional T2-based contrast (which underpins the use of the T2-weighted STIR sequence in clinical MRI) is lost. Ideally, it would be possible to take advantage of the T2-lengthening which underpins this ‘conventional’ MRI contrast in a quantitative fashion.

A relatively simple, quantitative option would be to perform conventional T2 relaxometry using a Carr-Purcell-Meiboom-Gill (CPMG) sequence. However, conventional T2-mapping techniques are likely to be heavily influenced by the large variations in fat content occurring in areas of inflammation, and may not specifically reflect changes in the relaxation time of the water resonance. T2-mapping could potentially be performed with fat suppression, but the quality of fat suppression is variable and likely to differ between different scanners, which might introduce bias into quantitative measurements.

An interesting alternative is to use a method which combines a Dixon-based water fat separation with a CPMG sequence, such that separate T2 measurements can be obtained for both fat and water (i.e. T2-fat, and T2-water) (209). In the ‘IDEAL-CPMG’ method, three echoes (rather than one) are collected between each 180° pulse; this amounts to one spin echo and two gradient echoes (see Figure 8-1below). This means that, for each triplet of echoes in the train, the water and fat signals can be separated. Having separated the signals, exponential fits can be performed individually for water and fat to derive T2-fat and T2-water.

Figure 8-1 - IDEAL-CPMG sequence as described by Janiczek et al. (209).
We have begun work towards implementing a similar approach on our scanner. Specifically, we are modifying a CPMG sequence on a Philips Ingenia system to enable variation in the position of the spin echo. By varying the position of the spin echo over a series of CPMG acquisitions (see below), we acquire echoes with different phase shifts allowing us to separate the fat and water signals around each spin echo. As with the IDEAL-CPMG, this enables calculation of T2-fat and T2-water measurements over the series of echoes in the CPMG train. The proposed approach is slightly different to IDEAL-CPMG (since the shifted echoes are acquired using interleaves of the CPMG acquisition, rather than in a single acquisition), but is likely to be simpler to implement and may also have some advantages in terms of the quality of fat-water decomposition.

![Proposed acquisition for T2-water measurement using interleaved CPMG trains with shifted echoes.](image)

Once the sequence modification has been completed, we plan to evaluate T2-water measurements in phantoms with varying fat content and T2 values, and in patients with spondyloarthritis. We will compare T2-water measurements with conventional T2 measurements to determine if they offer an advantage in terms of sensitivity or in terms of correlation/agreement with conventional scores. We expect that active inflammation (oedema) would cause an increase in T2-water. It is also possible that T2-fat measurements may be altered by inflammation (for example, in areas of fat metaplasia) and that this could be used as a marker of structural damage.

### 8.2.2 The Varying Composition of Normal Bone

An important outstanding problem is the varying composition of normal juxta-articular bone marrow. We have shown that ADC measurements are higher in
skeletally immature individuals compared to those who are skeletally mature (see Appendix B, Section 10.1), and it is likely that FF measurements vary in a similar fashion (FF measurements might be expected to increase as unossified cartilage is replaced by normal bone, which contains adipocytes). In adults, variations in marrow composition are likely to be relatively small compared to the effect of inflammation, but this variation can be substantial in paediatric and adolescent populations. Therefore, it will be essential to develop a method that can account for this variation.

One approach would be to make use of the fact that the composition of the skeleton at different anatomical locations is likely to be correlated. For example, in patients with unfused sacral apophyses, residual cartilage is present both in the interforaminal sacral bone (which is typically unaffected by inflammation) and in the subchondral bone (which becomes inflamed/oedematous in patients with sacroiliitis). It might be possible develop a regression model in which FF measurements from interforaminal bone are used to predict FF measurements and confidence intervals for juxta-articular bone. In this way, one could determine whether FF alterations are in line with the patient’s maturity, or whether they are likely to be due to a pathological process such as inflammation. An alternative approach would be to use statistical parametric mapping, whereby each patient’s images are registered to an atlas and abnormal regions are identified through a voxel-wise statistical comparison with that atlas (323).

### 8.2.3 Measuring Bone Formation and Destruction with MRI

Inhibiting spinal ankylosis is seen as a new ‘frontier’ in spondyloarthritis research, and is a currently a focus of development for pharmaceutical companies. There is some tenuous evidence that existing therapies (particularly NSAIDs) can inhibit new bone formation (324,325), but this remains controversial, and the evidence for TNFi’s reducing new bone formation is weak (25–27). As a result, new therapies – particularly inhibitors of IL-17A (an inflammatory cytokine which has a role in bone metabolism) such as secukinumab – are being developed with the express aim of reducing new bone formation (326). However, current approaches to measuring structural damage rely heavily on visual scoring of plain radiographs, which is again subjective and likely to be insensitive structural damage, but the radiation exposure is undesirable (particularly in young people). Therefore, an MRI method which
could reliably monitor bone formation would potentially be very valuable for drug
development, clinical trials and, ultimately, clinical practice.

The use of $R_2^*$ and quantitative susceptibility mapping has been discussed in
Chapters 5 and 7. Both techniques were able to detect BMD changes in phantoms,
but were significantly influenced by fat content in vivo, meaning that we could not
confidently identify changes in BMD in areas of fat metaplasia. It may be that the
lack of difference in susceptibility between these areas is a true biological
phenomenon, and it is possible that measurement of susceptibility in a different
location (for example in the joint itself) might be a more sensitive marker of bone
formation. Nonetheless, new bone formation is an important process in
spondyloarthritis, and there is clearly scope for optimization of the techniques.

Recently, a different approach using CSE-MRI to measure BMD has been proposed
by Ho et al., relying on the loss of signal created by increased BMD on synthesized
in-phase images (327). Using a relatively simple signal model, the estimated BMD
can be estimated based on the signal in the voxel, the maximum in-phase signal
intensities in the image (where there is assumed to be no mineral), and the signal
intensity from a hydroxyapatite phantom with known BMD (327). Despite the
requirement for the phantom, this method is attractive for its simplicity and could be
easily implemented on a clinical scanner.

BMD can also be measured using ultra-short echo time (UTE) or zero echo time
(ZTE) MRI, which acquire signal directly from mineralized bone (328). There has
been recent interest in using low-dose CT to evaluate structural damage (329), and
UTE/ZTE can generated CT-like images without the need for ionizing radiation
(330,331). Although these methods can be technically demanding, a number of
methods are now available and these may become easier to implement as the
technology improves (332).

It should be noted that BMD variations are an important pathological feature of
number of other diseases, including osteoporosis, multiple myeloma (where bone
loss at the site of focal ‘lytic’ lesions is a characteristic feature) and potentially
osteomyelitis, where infection causes bone destruction. Therefore, a reliable MR-
based method for measuring BMD could have a substantial clinical impact.
8.2.4 Whole Body Imaging for Measuring the Inflammatory Burden

One of the main advantages of CSE-MRI sequences is their speed compared to conventional MRI. Moreover, parallel imaging can potentially be used to accelerate these acquisitions with relatively little impact on FF measurements. Although this increase in speed could simply be used to reduce scan time, the alternative is to make use of it by imaging the whole body. WB-MRI could enable the detection and quantification of ‘occult’ inflammation at asymptomatic sites, which would not be possible using conventional MRI protocols.

Several groups have begun to investigate the use of qualitative whole body MRI for patients with spondyloarthritis, and for patients with JIA who may have involvement of multiple joints (333). By detecting early, ‘prestructural’ inflammatory involvement and treating accordingly, it is hoped that the subsequent structural damage can be limited or avoided altogether. Furthermore, around 50% of spinal lesions are thought to occur in the thoracic spine, which is included in WB-MRI protocols but may not be included in conventional clinical protocols (334). There is increasing interest in the role of whole-body MRI in multifocal aseptic osteitis in children and adolescents, where it may depict both axial and peripheral skeletal involvement and appears to outperform conventional techniques (bone scintigraphy and radiography) (335). Similarly, WB-MRI is becoming used more commonly in children and adults with chronic recurrent multifocal osteomyelitis (CRMO) and SAPHO syndrome, where it allows detection the detection of spine, pelvic and anterior thoracic wall inflammation in addition to peripheral involvement (333,336,337). There may also be a role for WB-MRI in systemic sclerosis, which causes multifocal involvement of subcutaneous tissues, muscles, fascia, synovium, tendons and entheses (338), and in imaging ischaemic lesions related to steroid use, lupus or haemoglobinopathies (333).

Going forward, applying quantitative imaging methods to WB-MRI could enable objective quantification of the inflammatory burden, and reduce the time needed for image reporting. However, achieving reliable, quantitative measurement of inflammation at multiple joints represents a major technical challenge. A simple first step would be to design specific ROI tools for each joint; however, this would be time consuming and might lead to substantial user-dependence in its own right. An alternative solution could be to segment the skeleton, thereby enabling the
Chapter 8 - Discussion and Perspectives

generation of ‘skeleton-only’ FF maps, whose composition would presumably change in patients with inflammation (as proof of principle, a clustering-based segmentation algorithm was used to segment a single subject’s sacrum, as shown in Figure 8-3).

This approach is loosely analogous to a method described by Blackledge et al., in which high b-value diffusion-weighted images are used to measure tumour volume in patients with cancer (339). The difficulty with inflammation is that the actual volume of disease may be relatively small; therefore, it may be challenging to separate the effects of inflammation from those of the normal skeleton. It might be possible to separate inflamed pixels from normal bone using clustering methods which identify the specific properties of inflamed pixels compared to normal pixels or, in the future, using machine learning methods which might also incorporate spatial information and prior probabilities based on other subjects. Optimisation of the inputs to such algorithms might help to improve their performance; for example, adding post-contrast CSE-MR images to the pre-contrast images (ideally both with the same field of view) may help to separate inflamed bone with a greater specificity. An advantage of CSE-MRI is that the same protocol could be used before and after contrast; a semi-quantitative ‘fractional enhancement’ parameter could then be derived from the water-only images from each acquisition.

Figure 8-3 - Example of segmentation using FF and R2* maps in a subject's sacrum. The images are courtesy of Dr Jorge Cardoso, who implemented the segmentation algorithm used to generate these images.
Moving forward, these problems will be best addressed by close collaborations between quantitative imaging researchers and computational imaging groups, allowing image analysis methods to be effectively tailored to the clinical need.

8.2.5 ‘Functional’ Imaging of Inflammation?

Although the techniques described in this thesis offer a number of advantages over existing methods, they still only measure inflammation in an indirect sense; they rely on the physical properties of the inflammatory exudate, but give us little information about the type of immune cells present or about the metabolism of the tissue. Future developments may involve attempting to characterize these ‘functional’ aspects of inflammation.

A number of preclinical groups have shown that is possible to label and track specific inflammatory cells (including dendritic cells and T-cells) in vivo (340,341). These methods have shown encouraging results in simple rodent models of inflammation (341), but have not been investigated in humans. Unfortunately, gaining ethical approval for this work is a major obstacle, and much more work is required before genuine clinical translation becomes a reality.

Similarly, hyperpolarized MRI can be used to detect metabolic alterations, and some groups have begun to investigate its potential as an inflammatory marker. Similar to tumours, inflammation causes an increased metabolic demand and often induces hypoxia; levels of pyruvate and lactic acid have both been shown to be elevated in the joints of patients with arthritis (342–346). Recently, Mackenzie et al. demonstrated the use of hyperpolarized MRI for imaging inflammation in a rat model of inflammatory arthritis (347). After injection of hyperpolarized $^{13}$C-pyruvate, they found an increase in the amount of $^{13}$C-lactate and in the lactate-to-pyruvate ratio at inflamed sites (347). They therefore argue that this method could be used to quantify inflammatory activity.

Hyperpolarized MRI could also be used to detect the pH derangements which are expected to accompany tissue hypoxia. For example, Gallagher et al. showed that $^{13}$C-labelled bicarbonate can be used to detect pH in vivo (348), whilst Duwel et al. have recently demonstrated that $^{13}$C-labelled zymonic acid can also be used as a pH marker (349). Although hyperpolarized MRI also remains some distance from clinical use, this technique has already been used in humans in several oncological
studies, which may shorten the path to implementation in patients with inflammatory diseases.

8.3 Towards Clinical Implementation

The work presented in this thesis predominantly consists of single-site studies aimed at technical and preliminary biological validation of the candidate biomarkers. However, if these biomarkers are to be widely used in clinical practice, it will ultimately be necessary to demonstrate that they lead to a net improvement in health outcomes, or provide useful information for diagnosis, treatment or management – in other words to show the clinical utility of the biomarkers (1,3). The roadmap described by O’Connor et al. provides a useful framework for planning future validation and qualification studies (3); the major steps which will need to be undertaken for inflammatory biomarkers are considered here.

Under the ‘biomarker roadmap’ framework, the process of technical validation for CSE-MRI in inflammation has begun but has not yet been completed. We have assessed repeatability at a single site, and reproducibility across three different scanners (one from each of the major vendors). We also plan to perform a larger study taking phantom data from twenty different sites, which will help to estimate reproducibility more precisely. Both of these multi-site studies have also begun to develop infrastructure that can be used for subsequent clinical studies using QIBs of inflammation in spondyloarthritis.

In order to standardize the process of QIB measurement, all of the necessary analyses for this larger study will be performed at one centre (i.e. by our group at UCLH), thereby optimizing the likelihood of a validated QIB being qualified as fit-for-purpose for clinical use (3). To facilitate this study and future research, we are developing a software ‘pipeline’ which will be used to process all of the image data from these sites, designed to minimize variability in QIB measurements due to differences in system software and hardware. This pipeline will accept CSE-MRI data in essentially any form, including pre-processed images from vendor-supplied quantitative packages (the most widely used being mDixon Quant, IDEAL-IQ and Liver Lab from Philips, GE and Siemens), simpler two- or three-point Dixon methods, or using raw complex multi-echo data. Images from all centres will undergo a rigorous quality assurance process using a standardized checklist of
acceptable acquisition parameters. Further, all centres will be asked to scan a standardized fat-water phantom (which will be manufactured within our group), enabling specific assessments of the bias in PDFF measurements at each site. To enable standardized ROI placement on the PDFF maps, the software pipeline will incorporate the measurement tool described in Chapter 6.

In terms of biological validation, although ADC measurements have been shown to decrease in patients undergoing biologic therapy, the same effect has not yet been demonstrated for PDFF. This could be assessed by scanning patients before and after commencement of biologic therapy, ideally in a prospective fashion. An additional form of biological validation would be to compare imaging measurements against histology. Due to ethical considerations, it is relatively difficult to obtain histological samples from patients with spondyloarthritis, but some specimens can be obtained in patients undergoing joint replacement surgery. If these patients could be identified in advance, imaging-pathology correlation would be ideally established using 3D histology [PDFF and R2* maps have successfully been correlated with histology in carotid plaques (350)]. 3D histology is important because, as with cancer imaging, spatial heterogeneity is common at inflamed sites and is unlikely to be accurately represented by single sections.

Clinical validation (qualification) of QIBs typically occurs relatively late in the development pathway and is often achieved using multicentre clinical trials (351). Again, imaging biomarker roadmaps provide a useful guide regarding potential endpoints for these studies, although these endpoints are likely to differ compared to QIB studies performed in cancer (351). In inflammatory diseases, the ultimate goal is to qualify the QIBs as prognostic for quality of life (in terms of pain, disability, and social/employment activities) or remission after treatment. Importantly, these multicentre trials will need to be powered adequately to demonstrate clinically useful effects, and rigorous and detailed statistical reporting standards will be essential.

In the case of spondyloarthritis, it would be interesting to compare the predictive capability of QIBs with standard visual scoring. These visual scoring methods are not widely used in clinical practice so the QIBs would not necessarily need to outperform them in order to demonstrate utility; however, an averaged visual score from several expert observers could act as a useful reference standard.
The ultimate ‘gold-standard’ in terms of clinical qualification would be to demonstrate improved outcomes in patients undergoing qMRI compared to those who did not. However, this is currently a distant prospect and would likely require very large sample sizes in order to achieve sufficient power.

8.4 Conclusion

Quantitative imaging of inflammation is in its infancy. Only a small number of groups have used quantitative MRI to image inflammation, and more work is needed before these techniques can be considered as mature research tools, let alone clinical tools. However, there is an undeniable need for a simple, objective method for quantifying inflammation – both in spondyloarthritis and other inflammatory diseases - and qMRI is well-suited to the task. In the years to come, further developments in imaging technology are likely to enable more sophisticated evaluation of the inflammatory process, of the effects of inflammation on the architecture of the bone, and of metabolic processes occurring at inflamed sites. The ultimate goal will be to translate these research methods into simple, clinically-useful tools which can guide therapeutic decision making in the clinic. Successful translation will ultimately require large, multi-site validation studies, and will likely depend on close collaborations between imaging and rheumatological communities.
9 Appendix A

9.1 Physical Basis of MR Signal

9.1.1 Conversion between laboratory and rotating Frames

The transformation between the lab frame and the rotating frame can be represented using a rotation matrix. If $m_x$, $m_y$ and $m_z$ are the three components of the magnetization in the rotating frame, and $m_{x'}$, $m_{y'}$ and $m_{z'}$ are the components in the laboratory frame, then

$$
\begin{bmatrix}
m_x \\
m_y \\
m_z
\end{bmatrix} =
\begin{bmatrix}
\cos(\omega_0 t) & -\sin(\omega_0 t) & 0 \\
\sin(\omega_0 t) & \cos(\omega_0 t) & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
m_{x'} \\
m_{y'} \\
m_{z'}
\end{bmatrix}.
$$

Similarly, the components in the laboratory frame can be determined from the components in the rotating frame

$$
\begin{bmatrix}
m_{x'} \\
m_{y'} \\
m_{z'}
\end{bmatrix} =
\begin{bmatrix}
\cos(\omega_0 t) & \sin(\omega_0 t) & 0 \\
-\sin(\omega_0 t) & \cos(\omega_0 t) & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
m_x \\
m_y \\
m_z
\end{bmatrix}.
$$

The relationship between the lab and rotating frames can also be given in terms of the rates of change of the vectors in the two frames:

$$
\left(\frac{d\bar{m}(t)}{dt}\right)_{lab} = \left(\frac{d\bar{m}(t)}{dt}\right)_{rot} + \vec{\Omega} \times \bar{m}(t)
$$

where 'lab' and 'rot' subscripts indicate the laboratory and rotating frames and $\vec{\Omega}$ is an angular velocity vector ($\vec{\Omega} = -\omega_0 \hat{z}$) aligned parallel to the $B_0$ field (352).

9.1.2 MR signal during frequency encoding

Mathematically, the total signal $S$ at a given time $t$ during the frequency encode gradient can be written

$$
S(t) = \int p(x)e^{ixkFE} \, dx.
$$
where \( \rho(x) \) is the proton density at a given spatial location, \( t \) is the duration of the gradient (relaxation has been ignored for the sake of simplicity). The constant \( k_{FE} \) corresponds to the specific spatial frequency being interrogated at time \( t \), which is dictated by the strength of the gradient and the time for which it has been applied, such that

\[
k_{FE} = \gamma G_{FE} \Delta t m
\]

\( m \) represents the number of the sample, and \( \Delta t \) represents in the interval between samples. The correspondence between the signal and the value for \( k_{FE} \) gives rise to the storage of signal data in a format known as k-space, which is described in Section 2.3.3.2. The size of the signal for each value of \( k_{FE} \) can be directly recorded in k-space, such that each frequency encode (or readout) gradient fills one row.

### 9.1.3 MR signal during phase encoding

The total signal from the sample is given by

\[
S(t) = \int \int \rho(x,y)e^{ixk_{FE}y}e^{yk_{PE}t} \, dx \, dy.
\]

Similar to \( k_{FE} \), the constant \( k_{PE} \) is given by

\[
k_{PE} = \gamma \Delta G_{PE} nt
\]

where \( \Delta G_{PE} \) is the size of the increment of the phase encoding gradients, and \( n \) is the number of phase encoding steps (or the number of times the pulse sequence is repeated).

### 9.2 CSE-MRI Algorithms

The following sections contain code which was written by this author to implement two previously-described CSE-MRI algorithms.

#### 9.2.1 IDEAL

The following code was written based on the description of the IDEAL algorithm given in Yu et al. (204).
%Assumes single spectral peak of fat
%Incorporates T2* decay
%Based on Yu et al. JMRI 26:1153-1161(2007)

%For single slice (20)
sl=20;
sliceimage=imDataParams.images(:,:,sl,:,:);

%Specify Echotimes (ms)
t = imDataParams.TE;

%Specify fat-water shift (kHz)
deltaF=0.45;

%imSingleSlice.FieldStrength = 3;
%imSingleSlice.PrecessionIsClockwise =0;

%%%%%%%%%%%%%%%%%
-----------
%IDEAL
-----------

%First, create known matrices

%Create matrix A (known)
A(1:6,1)=1;
for k=1:6
    A(k,2)=exp(i*2*pi*deltaF*t(1,k));
end

%For single slice
for x=1:320
    for y=1:320
        %Get data for individual pixel
        for k=1:6
            Signal(k,1)=sliceimage(y,x,1,1,k);
        end
        %Specify initial guess of psihat
        %psihat(y,x)=0.1 + 0.1*i;
        psihat(y,x)=0;
        %Create matrix P(psihat) based on initial guess of psihat
        for k=1:6
            P(k,k)=exp(i*2*pi*psihat(y,x)*t(1,k));
        end
        %Create matrix (P-psihat) based on initial guess of psihat
        %Pminus=(eye(6)/P) (this was the previous implementation)
        for k=1:6
            Pminus(k,k)=exp(i*2*pi*(-psihat(y,x))*t(1,k));
        end
        %Initial water/fat estimation (find p)
p=((inv(A.'*A))*A.'*Pminus)*Signal;
        %Now find error terms and iterate (very important to start this loop after
%initial guesses.. otherwise these will constantly be replaced) %%%%%%%%%

it=15;
for l=1:it

%Create matrix B(w,f)
for k=1:6
    B(k,1)=(w+f*exp(i*2*pi*deltaF*t(1,k)))*i*2*pi*t(1,k);
    B(k,3)= exp(i*2*pi*deltaF*t(1,k));
end
B(1:6,2)=1;

%Derive error terms
Error = ((inv(B.'*B))*B.')(Pminus*Signal - A*p);

%Update complex field map
psihat(y,x)=psihat(y,x)+Error(1,1);

%Now repeat
%Old code for updating f and w - wrong?
%p(1,1)=p(1,1)+Error(2,1);
%p(2,1)=p(2,1)+Error(3,1);
%  w=p(1,1);
%  f=p(2,1);
end

%Find final parameters
Water(y,x)=w;
Fat(y,x)=f;
%round(abs(f),3)
FF(y,x)=Fat(y,x)/(Water(y,x)+Fat(y,x));
Psi(y,x)=abs(real(psihat(y,x))); %Phase
R2star(y,x)=round(abs(imag(psihat(y,x))*2*pi),3); %R2*
end

9.2.2 Magnitude-based fitting

The following code was based on the description of magnitude-based fitting given by Bydder et al. (218). In the implementation shown here, the algorithm was modified such that the start points for the fit were determined based on preliminary fat and water images (labelled ‘Fatguess’ and ‘Waterguess’) derived from the Berglund method (198) from the fat water-toolbox (186). The manufacturer’s own Dixon images can also be used to provide these guesses (see Section 4.3.2.4).

%Magnitude refine
%Uses multipeak fat model with lsqcurvefit (spectrum based on peanut oil,
%Yu Shimakawa 2008) - this spectrum is almost identical to human subcutaneous fat

%Modular code: run A1, A2, A3 in succession; then B1 OR B2. C1, C2... can be %used to refine the initial maps.

slice=28
TE=1000*imDataParams.TE;

%Get magnituide images
for echoN=1:6
 M(:, :, echoN)=abs(imDataParams.images(:, :, slice, 1, echoN));
end

%Pre-declare fitting options (better outside loop)
ft = fittype( 'abs((F*0.62*exp((i*2.64)*x)+F*0.15*exp((i*2.00)*x)+F*0.10*exp((i*0.59))*x)+F*0.06*exp((i*2.97)*x)+F*0.03*exp((i*1.47)*x)+F*0.04*exp((i*0.29)*x)+W)*exp(-v*x)'' , 'independent', 'x', 'dependent', 'y' );

%M Main fat peak at 3.5ppm from water: for w in ms, w=3.5*128*2*pi/1000=2.81
%abs((F*1*exp((i*2.81)*x)+W)*exp(-v*x))

%Determine fat and water guesses
Fatguess=abs(F.amps(:, :, slice));
Waterguess=abs(W.amps(:, :, slice));

%Prefill
Sf=zeros(320,320);
Sw=zeros(320,320);
v=zeros(320,320);
Rsquare=zeros(320,320);
a=zeros(1,3);

for posX=1:320
 %Describes location of chosen pixel
 parfor posY=1:320
 %Get pixel data for single pixel
 Mag=M(posY,posX,:);
 Mag=reshape(Mag,1,6);
 [xData, yData] = prepareCurveData( TE, Mag );

 % Set up fittype and options.
 opts = fitoptions( 'Method', 'NonlinearLeastSquares' );
 opts.Display = 'Off';
 opts.MaxFunEvals=100
 opts.MaxIter=100
 opts.Lower = [0 0 0];
 opts.Upper = [2000 2000 0.3];
 opts.StartPoint = [Fatguess(posY,posX) Waterguess(posY,posX) 0.1];
 %Sf, Sw v

 % Fit model to data.
 [fitresult, gof] = fit( xData, yData, ft, opts );
 a=coeffvalues(fitresult);
 Sf(posY,posX)=a(1,1);
 Sw(posY,posX)=a(1,2);
 v(posY,posX)=a(1,3);
% Add Rsquare map
Rsquare(posY,posX)=gof.rsquare;

end
end

FFmag2=Sf./(Sf+Sw);
figure(1), imshow(FFmag2)
colormap('default');
colorbar;
title('Fat fraction')

figure(2), imshow(v,[])
colormap('bone');
colorbar;
title('R2*')

figure(3), imshow(Rsquare,[])
colormap('hot');
colorbar
caxis([-1, 1])
title('Rsquared')

9.2.3 ROI propagation for BEACH Tool

9.2.3.1 Createpoly

% Function createpoly polygonal ROI encompassing the whole joint
function [xy1,xy2] = createpoly(polypoints,width,im)

% Preallocate coords
x1=[];
y1=[];
x2=[];
y2=[];

% Define flaring factor
fl=1.2;
% Define spacing factor (determines how far apart the lines are)
sp=15;

for l=2:(0.5*numel(polypoints)-3) % Range of l determines which drawn
lines have perplines (The penultimate line doesn't have any perplines
- these are provided by the last line)
% Get coords and length for each line
coords(:,1)=polypoints(l:l+1,1);
coords(:,2)=polypoints(l:l+1,2);

linelength=sqrt(((coords(1,1)-coords(2,1))^2+(coords(1,2)-
(coords(2,2))^2)));

% Determine number of intersections needed
line=improfile(im,coords(:,1),coords(:,2));
umintersections=round(numel(line)/sp);
umintersections(numintersections<1)=1; % Minimum is one intersection
at end of line

% Generate line for each intersection
for k=1:numintersections
xcoord(k,1)=coords(1,1) + ((coords(2,1) - coords(1,1))*(k/numintersections));
ycoord(k,1)=coords(1,2) + ((coords(2,2) - coords(1,2))*(k/numintersections));

grad(k,1)=(coords(2,2)-coords(1,2))/(coords(2,1)-coords(1,1));
perpgrad(k,1)=-1/grad(k,1);

xlength(k,1)=(width^2/(1+perpgrad(k,1)^2))^0.5; %Gives length in x and y direction of perplines
ylength(k,1)=perpgrad(k,1)*xlength(k,1);

%Record x and y coords of all perplines
x1(numel(x1)+1,1)=xcoord(k,1)+xlength(k,1);
y1(numel(y1)+1,1)=ycoord(k,1)+ylength(k,1);
x2(numel(x2)+1,1)=xcoord(k,1)-xlength(k,1);
y2(numel(y2)+1,1)=ycoord(k,1)-ylength(k,1);

x=[x1(numel(x1),1) x2(numel(x2),1)]; %Chooses current (last) x1 value
y=[y1(numel(y1),1) y2(numel(y2),1)];
end
end

%Generate start and end lines
%For first line
grad1=(polypoints(2,2)-polypoints(1,2))/(polypoints(2,1)-polypoints(1,1));
perpgradl=-1/grad1;
xlen=((fl*width)^2/(1+perpgradl^2))^0.5; %Gives length in x and y direction of perplines %fl is flaring factor
ylen=perpgradl*xlen;

fx1=polypoints(2,1)+xlen;
fy1=polypoints(2,2)+ylen;
fx2=polypoints(2,1)-xlen;
fy2=polypoints(2,2)-ylen;

fx=[fx1 fx2];
y=[fy1 fy2];

%For last line
n=numel(polypoints)*0.5;

gradl=(polypoints(n,2)-polypoints(n-1,2))/(polypoints(n,1)-polypoints(n-1,1));
perpgradl=-1/gradl;
xlen=((fl*width)^2/(1+perpgradl^2))^0.5; %Gives length in x and y direction of perplines
ylen=perpgradl*xlen;

lx1=polypoints(n-1,1)+xlen;
ly1=polypoints(n-1,2)+ylen;

lx2=polypoints(n-1,1)-xlen;
ly2=polypoints(n-1,2)-ylen;

lx=[lx1 lx2];
ly=[ly1 ly2];

end
%Combine coordinate data - creates x and y columns for each row of
%perpoints
%xy1 is one side of the polygon ('left')
xy1(1,1)=fx1;
xy1(1,2)=fy1;
xy1(2:numel(x1)+1,1)=x1;
xy1(2:numel(y1)+1,2)=y1;
xy1(numel(x1)+2,1)=lx1;
xy1(numel(y1)+2,2)=ly1;

xy2(1,1)=fx2;
xy2(1,2)=fy2;
xy2(2:numel(x2)+1,1)=x2;
xy2(2:numel(y2)+1,2)=y2;
xy2(numel(x2)+2,1)=lx2;
xy2(numel(y2)+2,2)=ly2;

%xy2 is the other side ('right')
xy2=fliplr(xy2); %inverts xy2

9.2.3.2 Use of createpoly to define two polygonal ROIs

case 'impoly'

hroi = impoly(vin{:},'Closed',false) ;
setColor(hroi,'red');
posn = getPosition(hroi) ;
if size(posn,1) < 3
 warning([ 'Less than 3 points in impoly'])
 delete(hroi)
 hroi = draw_imroi(handles, varargin{:}) ;
end

get(handles.edit5)
width=handles.edit5.Value
get(handles.edit8)
widthinner=handles.edit8.Value
im=imgcf
im=double(im)
polypoints=hroi.getPosition
delete(hroi)

%Get wide and narrow polygons
[xy1wide,xy2wide]=createpoly(polypoints,width,im);
[xy1narrow,xy2narrow]=createpoly(polypoints,widthinner,im);

%Create left and right polygons
xyL=[xy1wide; flipud(xy1narrow)];
xyR=[xy2wide; flipud(xy2narrow)];

hroi=impoly(gca,xyL,'Closed',true);
set(hroi,'Tag', 'impoly');
hroi2=impoly(gca,xyR,'Closed',true);
set(hroi2,'Tag', 'impoly');
10 Appendix B

10.1 Association of the Apparent Diffusion Coefficient with Maturity in Adolescent Sacroiliac Joints

10.1.1 Introduction

Sacral ossification begins in the first two sacral segments in utero (353,354). As the sacral apophyses ossify during childhood, residual cartilaginous connections which link the SIJs and the neural foramina gradually disappear. The apophyses typically fuse between the ages of 16 and 20, although sometimes the SIJs remain immature well into late adolescence (353,355). Immature sacroiliac joint morphology may be misinterpreted as inflammation because unossified cartilage causes juxta-articular areas of high signal on STIR images. To our knowledge, there are no previous studies investigating the influence of joint maturity on apparent diffusion coefficient (ADC) values as measured using DWI.

We hypothesised that ADC values would vary with joint maturity, since unossified bone would be expected to contain a higher proportion of water than fully mineralised bone, and there is a well-known relationship between water content and ADC (178). Therefore, we evaluated the association between SIJ maturity and both ADC and nADC measurements in adolescent and young adult patients with non-inflammatory back pain.

10.1.2 Methods

This study was covered by IRB approval (REC ref: 11/LO/0330) and informed consent was waived due to its retrospective nature.

Subjects

A picture archiving and communication system (PACS) search was used to identify 74 adolescent and young adult patients (aged 12-24 years) who had an MRI of the sacroiliac joints performed at our institution from January 2010 to June 2015 without evidence of sacroiliitis. For all potential subjects, the electronic medical record was reviewed to ensure a final clinical diagnosis of mechanical, non-inflammatory back pain and normal inflammatory markers (defined as a serum C-reactive protein level less than 5mg/L, and an erythrocyte sedimentation rate less than 7mm/hr). Individuals with inflammatory arthritis or connective tissue disease...
were excluded (n=17). Patients whose images were substantially degraded by artifact (particularly fat ghosting due to inadequate fat suppression) were also excluded (n=2).

**MRI technique**

Both conventional and diffusion-weighted images were acquired on a 1.5T scanner (Avanto; Siemens, Germany), using the imaging protocol described in Chapter 3.

**Classification according to Maturity**

Axial and coronal T1-weighted images and coronal STIR images were reviewed in consensus by two observers (MHC and TB, with over twenty years and three years of musculoskeletal MRI experience respectively) to determine the degree of maturity of the SIJs. The joints were assessed at S1/2 and S2/3. Based on previous work describing age-related differences in the degree of fusion of the segmental apophyses (355), subjects were classified as either 'unfused', 'partially fused' or 'fused' as follows:

Unfused: The apophyses between the sacral segments are unfused (open), and there is a complete cartilaginous connection between the SIJs and neural foramina (Figure 10-1a,b). There are no areas of bony fusion visible on either the coronal or sagittal images. On the STIR images (Figure 10-1a), these patients often showed high signal bands of unossified bone adjacent to the SIJ on the sacral side, in continuity with the unossified intersegmental bone, although this was not used as a classification criterion.

Partially fused: There are some areas of fusion (i.e. there is an area of continuous bone joining the sacral segments) but the fusion is incomplete (Figure 10-1c). This group includes a spectrum of patients ranging from those with very early fusion, to those where fusion was almost complete (i.e. there are small areas of residual unossified cartilage).

Fused: The intersegmental apophyses are fully fused. There is no residual unossified cartilage. The contours of the sacroiliac joint are sharply defined (Figure 10-1d,e).
**ADC Measurement**

Figure 10-1 - Maturity classification system. In patients with unfused intersegmental apophyses (a,b), there are persistent cartilaginous connections between the joint and neural foramina (arrowhead), and high signal bands of unossified bone adjacent to the SIJ (arrow). In the ‘partially fused’ group (c), there is partial ossification of the apophyses seen medially (arrowedhead) but a thin subchondral band of high signal (arrow) remains. In the fused group (d,e), the joint margin is clearly defined (arrow) and the cartilaginous bands linking the neural foramina and SIJs have disappeared.
The ADC maps were analysed using the linear ROI method described in Chapter 3. Measurements were performed by two radiology registrars (TB and JR) with three and two years of musculoskeletal MR experience respectively, who were each blinded to the other observer’s ROI placement and ADC measurements, and to the maturity classification.

The mean of the two individuals’ mean joint values was taken as the ‘uncorrected’ ADC value. Additionally, a further ‘reference’ ROI was placed on normal sacral bone by each observer to enable calculation of nADC values. As before, the nADC value for each subject was defined as the ratio between the mean ADC of all joint line profiles and the mean ADC of the reference ROI from normal sacral bone.

ROIs were not placed on slices where the images were significantly degraded by fat ghosting artifact (these were typically visible as dark bands in the image, due to unsuppressed fat in combination with the EPI readout).

**Statistical analysis**

Mean ADC, nADC and reference ADC values for each subject were compared between the three groups using balanced one-way analysis of variance (ANOVA). A post-hoc multiple comparison test (Tukey’s honest significant difference criterion) was used to determine whether there were significant differences between groups. Additionally, we used a multilevel mixed-effects linear regression analysis to compare ADC, nADC and reference ADC values from individual ROIs between the three groups. In this analysis, ‘Maturity’ was used as he predictor variable, ADC, nADC or reference ADC was used as the outcome variable, and ‘Observer’ and ‘Subject number’ were used as grouping variables. The ‘Unfused’ group was used as the baseline for comparison. The ages of the three groups were compared using a one-way ANOVA, and sex was compared between the three groups using a 2x3 Fisher’s exact test.

Interobserver variability was assessed for ADC, nADC and reference ADC values using Bland-Altman 95% limits of agreement and the intraclass correlation coefficient (absolute agreement).
10.1.3 Results

Demographics

Fifty-five subjects were included in the study, of whom 36 were female and 19 were male, with a mean age of 15y 11m (range 10y2m to 18y11m).

Demographics for the three groups are summarised in Table 11. Patients in the unfused group were significantly younger than those in the fused group (p=0.011) and also younger than those in the partial group (p=0.051). There was no significant difference in the ages of subjects in the partial and fused groups (p=0.53), and no significant association between sex and fusion class (p=0.22, Fisher’s exact test).

<table>
<thead>
<tr>
<th></th>
<th>Fused</th>
<th>Partially fused</th>
<th>Unfused</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>14</td>
<td>29</td>
<td>12</td>
</tr>
<tr>
<td>Males (%)</td>
<td>6 (43%)</td>
<td>7 (24%)</td>
<td>6 (50%)</td>
</tr>
<tr>
<td>Mean age (SD)</td>
<td>16y 7m (1y 6m)</td>
<td>16y 1m (1y 4m)</td>
<td>14y 10m (1y 11m)</td>
</tr>
<tr>
<td>Age range</td>
<td>5y 4m</td>
<td>5y 7m</td>
<td>6y 10m</td>
</tr>
<tr>
<td></td>
<td>(13y 8m to 18y</td>
<td>(12y 7m to 18y 2m)</td>
<td>(10y 2m to 16y 11m)</td>
</tr>
</tbody>
</table>

Table 11 – Demographics for MM patients.

Comparison of Fused, Partially Fused and Unfused Groups

ADC, nADC and reference ADC values for the three groups (fused, partially fused and unfused) are shown in Figure 10-2.

i. ADC: The mean ADC values were 690 ± 174 x 10^-6 mm²/s in the fused group, 720 ± 156 x 10^-6 mm²/s in the partial group, and 842 ± 145 x 10^-6 mm²/s in the unfused group (Figure 10-2a). Mean ADC values were significantly higher in the unfused group than in the fused group (p=0.046). Mean ADC values were also lower in the unfused group than in the partially fused group (p=0.074) but there was no significant difference between partially fused and fused groups (p=0.82).

Accordingly, multilevel regression analysis found that individual ADC values in the unfused group were significantly higher than those in the fused group (p=0.019,
regression coefficient: \(-133 \times 10^{-6} \text{mm}^2/\text{s}, 95\% \text{ CI: -233 to -34}\) or the partial group (p=0.009, regression coefficient: \(-124 \times 10^{-6} \text{mm}^2/\text{s}, 95\% \text{ CI -227 to -20}\).

ii. Normalised ADC (nADC): The mean nADC values were \(1.23 \pm 0.14\) in the fused group, \(1.34 \pm 0.35\) in the partial group and \(1.40 \pm 0.27\) in the unfused group (Figure 10-2b). The difference between fused and unfused groups showed a trend towards statistical significance (p=0.093). There was no significance difference between partial and fused groups (p=0.64) or between partial and unfused groups (p=0.26).

iii. Reference ADC: Reference ADC values were \(571 \pm 140 \times 10^{-6} \text{mm}^2/\text{s}\) in the fused group, \(577 \pm 132 \times 10^{-6} \text{mm}^2/\text{s}\) in the partial group, and \(592 \pm 131 \times 10^{-6} \text{mm}^2/\text{s}\) in the unfused group (Figure 10-2c). There was no significant difference in reference ADC between any of the three groups (fused compared to partial: p=0.99, fused compared to unfused: p=0.91, partial compared to unfused: p=0.94).

Figure 10-2 - ADC, nADC and reference ADC measurements compared by skeletal maturity. ADC measurements were significantly higher in patients with unfused sacral apophyses than those with fused apophyses (p=0.046).
Inter-observer variability

For the ADC analysis, the Bland-Altman 95% limits of agreement were ± 47 across a range of values from 401 to 1166. The intraclass correlation coefficient was 0.99. The coefficient of variance was 0.23.

For the nADC analysis, the Bland-Altman 95% limits of agreement were ± 0.58 across a range of values from 0.93 to 2.41. The intraclass correlation coefficient was 0.62. The coefficient of variance was 0.23.

For the reference ADC, the Bland-Altman 95% limits of agreement were ± 160 across a range of values from 340 to 579. The intraclass correlation coefficient was 0.84. The coefficient of variance was 0.22.

10.1.4 Discussion

The sacroiliac joints undergo substantial structural changes during adolescence (353,355) and SIJ ADC values might therefore be expected to vary according to skeletal maturity. We found that SIJ ADC values were higher in patients with unfused sacral segmental apophyses compared to those with fused segmental apophyses. To our knowledge, this is the first report describing an association between skeletal maturity and SIJ ADC measurements.

ADC values in unfused subjects might be higher than in fused and partially fused subjects because the proportion of unmineralised bone adjacent to the joint is higher. Immature bone consists of cartilage and unossified or partially ossified bone, and would therefore be expected to contain a higher proportion of water than fully mineralised bone (356). ADC measurements have been shown to correlate with water content and collagen matrix structure in cartilage (357–359) and variations in water content are related to the degree of mineralization of the bone matrix (360). Age- and subject-related variations in the composition of bone marrow adjacent to the joint may also influence the measured ADC values, since red marrow displays significantly higher apparent diffusivity than yellow marrow (265,361).

Importantly, SIJ ADC values in unfused subjects may overlap with those previously reported in sacroilitis (243,247,286,362). For example, Vendhan et al. (363) report a mean joint ADC value of 1211 x 10^{-6} \text{mm}^2\text{s}^{-1} in ERA cases with sacroilitis, which is only slightly higher than the upper end of the ‘unfused’ normal range reported in
this work (range 687 to 1166 x 10^{-6} mm^2s^{-1}). In adults, reported ADC values in sacroiliitis vary widely from 480 x 10^{-6} mm^2s^{-1} (286) to 1310 x 10^{-6} mm^2s^{-1} (243), while focal areas of bone marrow oedema in adolescents with chronic nonbacterial osteomyelitis have been measured at 1600 x 10^{-6} mm^2s^{-1} (364). These results provide a clear indication that ADC values in adolescent joints cannot be viewed in isolation and must be interpreted in light of skeletal maturity.

This work provides new information regarding the reproducibility of ADC values in normal adolescent SIJs. In our study, the reproducibility of ADC values was excellent. Interestingly, the use of a reference ADC to normalise the data reduced the intraclass correlation coefficient and widened the Bland-Altman limits of agreement. This suggests that the use of uncorrected ADC values may provide better interobserver reproducibility than nADC, and may explain why the difference between fused and unfused groups was significant for ADC measurements and nonsignificant for nADC measurements. However, the use of a reference ADC may help to minimise scan variability that may occur due to the use of different scanning platforms when serial scans are acquired during patient treatment. We have been unable to assess this in the current study as these patients had a single scan only.

A limitation of the current study is that joint maturation is a continuous process, and classification of borderline patients into discrete groups can be difficult. Nonetheless, there is a very clear distinction between the fused and unfused groups we describe. The group of patients with partial fusion of the segmental apophyses is more heterogeneous, and includes patients whose apophyses have only just started to fuse or are almost fully fused. Additionally, these patients were only scanned at a single time point – it was not possible to observe a progression in ADC changes over time in individual patients. In this study, we only assessed apophyseal fusion at S1/2 and S2/3 since the apophyses below these levels are smaller – as a result, we could not consistently distinguish between partially fused and fused/unfused apophyses at S3/4 or S4/5 due to partial volume effects.

Another limitation of the present study is its retrospective nature. Ideally, one could prospectively recruit equal numbers of patients for the three groups (fused, partial and unfused) although this would be rather impractical because an MRI scan is required for classification purposes. Furthermore, the scans were only acquired on a
single scanner at one institution. It would be desirable to repeat this study prospectively in a larger cohort, ideally using multiple imaging platforms.

Further work will be required to develop strategies to allow for joint immaturity when quantitatively measuring inflammation of the sacroiliac joints. One approach would be to define a normal range of ADC values at different stages in sacral maturation. In JIA, it may be more practical to simply monitor joint ADC over time in individual patients since maturation would be expected to produce a gradual decrease in ADC whereas inflammation causes joint ADC to increase.

10.1.5 Acknowledgements


10.2 Diagnostic utility of whole body Dixon MRI in multiple myeloma – a multi-reader study

10.2.1 Introduction

In recent years, whole body-MRI (WB-MRI) has emerged as a valuable tool for assessing disease activity in multiple myeloma (MM)(365–369). MRI is a key component of the Durie-Salmon PLUS staging system (370), and the number of lesions identified on MRI correlates closely with mortality (371). As a result, WB-MRI is developing into a first-line imaging modality in MM (372,373).

The two major obstacles for widespread use of WB-MRI are cost and long scan times. It is therefore important to maximise diagnostic value but minimise acquisition time, particularly for MM patients who may be frail and in pain. To make best use of the available scan time, WB-MRI protocols typically include both anatomical imaging (for assessment of morphology, fractures and spinal cord compression (333,374)) and functional imaging (for assessing cellularity and perfusion (333,369,375,376)). However, imaging protocols vary substantially between centres: anatomical imaging may use T1-weighted or T2-weighted images (or a combination), and may implement spin echo- or gradient echo-based
When choosing sequences, considerations include image quality, acquisition time, the cost of data acquisition and storage, and interpretation time.

Recently, gradient echo-based Dixon MRI has been used for anatomical WB-MRI in MM, and has several advantages over conventional T1- or T2-weighted imaging. Dixon MRI enables the generation of four separate image types: in-phase (IP), out-of-phase (OP), water-only (WO) and fat-only (FO). Acquisition times are similar to those for conventional gradient echo imaging and shorter than for spin echo imaging. When reporting, the IP images can be viewed in a similar fashion to conventional T1-weighted images, whilst water and fat can be separately evaluated on WO and FO images.

However, it is uncertain whether the ‘additional’ images (OP, WO and FO) offer any additional diagnostic information compared to IP imaging alone, and if so which image type is optimal for reading. Therefore, the best approach to reporting WB-MRI is unclear: it is uncertain whether reviewing the IP images alone is sufficient, or whether the additional images provide additional information. Clarifying this issue could improve the accuracy of disease staging and also increase reporting efficiency - radiologists could begin their read by reviewing the most diagnostically-useful scans. Furthermore, reconstructing and storing the additional images would only be justified if they provided additional diagnostic information.

In this study, we aimed to evaluate radiologists’ diagnostic accuracy for detecting focal lesions on each of the four Dixon image types, using post-contrast and diffusion images as a reference standard. We hypothesised that sensitivity would be improved by using FO and WO images compared to IP images.

10.2.2 Methods

Subjects
This prospective study was performed with institutional review board approval (Research Ethics Committee reference 12/LO/0428). All patients gave written informed consent.

Thirty patients (13 males and 17 females, median age 55, age range 36-82) with clinically suspected symptomatic multiple myeloma were prospectively enrolled between June 2012 and September 2014. Patients were excluded if they had a
history of previous malignancy or previous chemotherapy/radiotherapy, estimated GFR < 50 mL/min/1.73 m², were unable to given informed consent or had a contraindication to MRI scanning. Further assessment showed that 26 out of 30 had MM, one had smoldering MM, two a had solitary plasmocytoma, and one had monoclonal gammopathy of uncertain significance. For each patient, clinical and biochemical parameters were recorded as shown in Table 12. Baseline interphase fluorescence in situ hybridisation (FISH) was performed on CD138-selected plasma cells from bone marrow samples, using probes for IGH translocations t(4;14), t(11;14) and t(14;16), del(17p), del(13) and 1p- /1q+ (379). Genetic risk was determined according to International Myeloma Working Group recommendations (380).

**Acquisition**

All subjects underwent WB-MRI imaging on a 3.0T wide-bore system (Ingenia; Phillips Healthcare, Best, Netherlands) using two anterior surface coils, a head coil and an integrated posterior coil. The WB-MRI protocol included coronal pre- and post-contrast modified Dixon (Dixon) acquisitions from which fat and water images and calculated in and out of phase images were reconstructed on the scanner using a two-point method (279) (TR 3.0ms, TE 1.02-18, flip angle 15°, slice thickness 5mm, bandwidth 1992 Hz/Px, acquisition matrix 196 x 238, SENSE factor 2, number of slice 120) in addition to diffusion and post-contrast imaging covering vertex to toe using ten contiguous anatomical stations (Table 13). The coronal images were ‘stitched’ together and presented as a head-to-toe whole body image to the reader; the images were then magnified according to the reader’s preference for specific analysis of the pelvis.
<table>
<thead>
<tr>
<th>Patient characteristics</th>
<th>Number or median (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>56 (36-80)</td>
</tr>
<tr>
<td>Chain isotype</td>
<td></td>
</tr>
<tr>
<td>IgG</td>
<td>17</td>
</tr>
<tr>
<td>IgA</td>
<td>5</td>
</tr>
<tr>
<td>Light chain</td>
<td>4</td>
</tr>
<tr>
<td>MGUS</td>
<td>1</td>
</tr>
<tr>
<td>Solitary plasmacytoma</td>
<td>2</td>
</tr>
<tr>
<td>Smoldering MM</td>
<td>1</td>
</tr>
<tr>
<td>ISS stage</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>13</td>
</tr>
<tr>
<td>II</td>
<td>13</td>
</tr>
<tr>
<td>III</td>
<td>4</td>
</tr>
<tr>
<td>Induction regiment</td>
<td></td>
</tr>
<tr>
<td>PAD</td>
<td>18</td>
</tr>
<tr>
<td>CVD</td>
<td>3</td>
</tr>
<tr>
<td>VTD</td>
<td>5</td>
</tr>
<tr>
<td>MPV</td>
<td>2</td>
</tr>
<tr>
<td>Bone marrow percentage plasma cells</td>
<td>65 (0-90)</td>
</tr>
<tr>
<td>Beta-2 microglobulin (mg/l)</td>
<td>3.3 (1.3-11.3)</td>
</tr>
<tr>
<td>Albumin (g/l)</td>
<td>40 (30-53)</td>
</tr>
<tr>
<td>Creatinine</td>
<td>56 (77.5-105)</td>
</tr>
<tr>
<td>Genetic risk group</td>
<td></td>
</tr>
<tr>
<td>Low/Standard risk</td>
<td>17</td>
</tr>
<tr>
<td>High risk</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 12 - Patient demographics, disease parameters and treatment. ISS, international staging system; DS-PLUS, PAD, bortezomib, doxorubicin, dexamethasone; CVD, cyclophosphamide, bortezomib, dexamethasone; VTD, bortezomib, thalidomide, dexamethasone; MPV, melphalan, prednisolone, bortezomib.
<table>
<thead>
<tr>
<th>Sequence Parameters</th>
<th>Dixon (pre and post contrast)</th>
<th>DWI (b0, 100, 300, 1000 s/mm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imaging Plane</td>
<td>Coronal</td>
<td>Transverse</td>
</tr>
<tr>
<td>Sequence type</td>
<td>Gradient echo</td>
<td>Single-shot spin echo with echo planar readout</td>
</tr>
<tr>
<td>Echo time (ms)</td>
<td>1.02/1.8</td>
<td>71</td>
</tr>
<tr>
<td>Repetition time (ms)</td>
<td>3</td>
<td>6371</td>
</tr>
<tr>
<td>Field of View (mm x mm)</td>
<td>502 x 300</td>
<td>500 x 306</td>
</tr>
<tr>
<td>Voxel size (mm x mm)</td>
<td>2.1 x 2.1</td>
<td>4 x 4.2</td>
</tr>
<tr>
<td>Number of Slices</td>
<td>120</td>
<td>40</td>
</tr>
<tr>
<td>Slice Thickness (mm)</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Acquisition Matrix</td>
<td>144 x 238</td>
<td>124 x 72</td>
</tr>
<tr>
<td>ETL</td>
<td>2</td>
<td>39</td>
</tr>
<tr>
<td>Acceleration factor (SENSE)</td>
<td>2</td>
<td>2.5</td>
</tr>
<tr>
<td>Pixel Bandwidth (Hz)</td>
<td>1992</td>
<td>3369</td>
</tr>
<tr>
<td>Scan time (s)</td>
<td>17</td>
<td>152</td>
</tr>
</tbody>
</table>

Table 13 - Sequence parameters for WB-MRI protocol.

**Image Assessment**

The individual sets of pre-contrast Dixon images were randomised and read by four consultant radiologists, who each had between five and fifteen years of specialist expertise in oncological MR imaging. All readers were blinded to clinical data and diagnosis. On each image set, each radiologist was asked to count the number of myeloma lesions present in the bony pelvis (pubis, ischium, ilium and sacrum) and to label these lesions on the images (up to a maximum of 20). If the disease was diffuse or there were over 20 lesions, the patient was assigned a lesion count of 20. Additionally, the radiologists were asked to provide a confidence score based on their degree of certainty that there were myeloma lesions in the pelvis on a 4-point Likert scale (1-no lesions, 2-indeterminate lesions, 3-likely myeloma lesions, 4-very likely myeloma lesions). After scoring, each labelled lesion was compared to a reference standard consisting of diffusion-weighted, pre- and post-contrast Dixon
imaging, which had been evaluated by a further consultant radiologist with over 20 years of experience in myeloma and MR imaging. On the reference imaging, all lesions demonstrating abnormal marrow signal compared to background marrow (i.e. hypointense on IP and FO images, and hyperintense on WO images) and which showed contrast enhancement or restricted diffusion were assigned as myeloma lesions and labelled on the images. For the reference standard, no maximum lesion count was used (i.e. all lesions were labelled) to ensure that all lesions on the IP, OP, FO and WO Dixon images could be compared directly to a reference lesion. Using the reference standard imaging, we also recorded whether patients had focal or diffuse disease (the diffuse category included patients with focal-on-diffuse infiltration).

For each Dixon image set, we compared each lesion with the reference standard to determine the number of per-set true positive lesions (TP), false positive lesions (FP) (i.e. those that were incorrectly identified as lesions); and false negative lesions (FN) (these were the ‘reference-standard lesions’ which were not identified). For each Dixon image type (30 sets per type), we determined the mean per-set lesion count, sensitivity (TP/TP+FN), positive predictive value (TP/TP + FP) and mean confidence score.

**Design and Statistics**

A summary of the study design is given in Figure 10-3. To account for clustering within the data, for each lesion detection metric (lesion count, sensitivity, positive predictive value and mean confidence score), values were compared across the four Dixon image types using a multilevel mixed-effects linear regression model, performed using Stata [Stata IC Version 14.1, College Station, USA]. Image type (i.e. IP, OP, FO or WO) was used as the predictor variable, and the value of the specific lesion detection metric being analysed (i.e. lesion count, sensitivity, positive predictive value or mean confidence score) was used as the outcome variable. Data were clustered at the level of ‘subject’ (patient) and ‘observer’ (radiologist). This analysis was repeated for the subgroup of patients who had diffuse disease (as determined by the reference standard assessment), and for the subgroup of patients with focal disease.
Percent Contrast and Contrast-to-Noise Ratio

Percent contrast and contrast-to-noise ratio (CNR) were calculated using a previously described method (377). Specifically, in patients with at least three focal lesions greater than 3mm in diameter, circular regions of interest (ROIs) were placed on the three largest focal myeloma lesions, and three further ROIs were placed in areas of bone marrow without focal lesions in the sacrum and iliac bones.
Percent contrast was calculated as:

$$\text{Percent Contrast} = \frac{(S_a - S_b)}{(S_a + S_b)} \quad [1]$$

where $S_a$ is the mean signal intensity of myeloma lesions and $S_b$ is the background marrow signal intensity.

Similarly, CNR was calculated as:

$$\text{CNR} = \frac{|S_a - S_b|}{\sqrt{(S_{asd} + S_{bsd})/2}} \quad [2]$$

where $S_{asd}$ and $S_{bsd}$ are the mean within-ROI standard deviation values for myeloma lesions and background marrow respectively. A one-way analysis of variance (ANOVA) with a post-hoc Tukey Kramer multiple comparison test was used to compare percent contrast and CNR between image series.

10.2.3 Results

Four radiologists read four image series for each of 30 patients (120 image series per radiologist), and identified 610, 955, 549 and 734 lesions respectively compared to 1560 reference lesions. An example of a focal lesion, as shown on the four Dixon image types, is given in Figure 10-4. A summary of the mean lesion count, true positives, sensitivity, positive predictive value and confidence score for each of the four image types is given in Table 14; these values are also shown graphically in Figure 10-5. The results of the regression analysis including confidence intervals are also provided in Table 14.
Figure 10-4 - Examples of focal MM lesions. There is a large focal lesion in the right ischium (solid arrow) and a smaller lesion in the right ilium (dashed arrow); both lesions are shown on unenhanced Dixon images (IP, OP, WO and FO) and on the reference images (consisting of DWI and post-contrast). The smaller lesion is less conspicuous on the IP and OP images than on the FO and WO images. *DWI was acquired in the axial plane; the smaller of the two lesions (in the right ilium) is again marked with a dashed arrow.

Figure 10-5 - Lesion count, sensitivity, positive predictive value (PPV) and confidence for each Dixon image type. Individual observers are shown in colour (see legend), and the mean value across all four observers is shown in black. Error bars indicate the 95% confidence interval.
Table 14 - Results of regression analysis (n=30). Lesion count, true positives, sensitivity, positive predictive value and confidence were compared between the four image types, using the in phase images as the baseline.

Regression analyses used image type were used as the predictor variable, and lesion count/TP/sensitivity/confidence were used as the outcome variable. Mean values were calculated by the regression analysis, and were equal to means calculated manually from all patients and all four radiologists.
Lesion count and True Positives
The mean lesion counts for each image type (averaged over all patients and all four radiologists) were 5.2 for IP, 5.8 for OP, 7.0 for FO and 5.7 for WO. Significantly more lesions were identified on the FO images than on the IP images (p=0.006), but there was no significant difference between OP and IP images (p=0.364) or WO and IP images (p=0.504).

Of the identified lesions, the mean number of true positives was 4.9 for IP, 4.6 for OP, 6.5 for FO and 5.1 for WO. Significantly more true positive lesions were identified on the FO images than on the IP images (p=0.008), but there was no significant difference in true positives between OP and IP images (p=0.633) or WO and IP images (p=0.702).

Sensitivity and Positive Predictive Value
The mean sensitivity for each image type was 0.34 for IP, 0.32 for OP, 0.42 for FO and 0.36 for WO. Sensitivity was significantly higher on the FO images than on the IP images (p=0.023), but there was no significant difference between OP and IP images (p=0.696) or between WO and IP images (p=0.590).

The mean positive predictive values were 0.86 for IP, 0.67 for OP, 0.81 for FO and 0.82 for WO. There was no significant difference in PPV for FO compared to IP (p=0.146) or WO compared to IP (p=0.617). However, positive predictive values were significantly poorer on OP images than on IP images (p=0.000).

Confidence Score
The mean confidence scores were 2.48 for IP, 2.65 for OP, 2.73 for FO and 2.68 for WO (1-no lesions, 2-indeterminate lesions, 3-likely myeloma lesions, 4-very likely myeloma lesions). Confidence scores were higher on all three image types than on the IP images (OP compared to IP: p=0.063, FO compared to IP: p=0.006, WO compared to IP: p=0.033).

Sub-group analysis
Of 30 patients, there were 23 patients in the focal disease group (this included the two patients with solitary plasmacytoma) and six patients in the diffuse disease group. True positives, sensitivity and PPV for focal and diffuse groups are given in Table 15 and Table 16 respectively.
### True positives

<table>
<thead>
<tr>
<th>Image type</th>
<th>Mean</th>
<th>Difference in means (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>5.1</td>
<td>Baseline</td>
<td>-</td>
</tr>
<tr>
<td>OP</td>
<td>5.0</td>
<td>-0.08 (-1.38 to +1.21)</td>
<td>0.900</td>
</tr>
<tr>
<td>FO</td>
<td>6.8</td>
<td>+1.77 (+0.47 to +3.07)</td>
<td>0.008</td>
</tr>
<tr>
<td>WO</td>
<td>5.2</td>
<td>+0.11 (-1.19 to +1.42)</td>
<td>0.863</td>
</tr>
</tbody>
</table>

### Sensitivity

<table>
<thead>
<tr>
<th>Image type</th>
<th>Mean</th>
<th>Difference in means (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>0.37</td>
<td>Baseline</td>
<td>-</td>
</tr>
<tr>
<td>OP</td>
<td>0.36</td>
<td>-0.00 (-0.09 to +0.08)</td>
<td>0.976</td>
</tr>
<tr>
<td>FO</td>
<td>0.46</td>
<td>+0.09 (-0.01 to +0.18)</td>
<td>0.037</td>
</tr>
<tr>
<td>WO</td>
<td>0.38</td>
<td>+0.02 (-0.07 to +0.10)</td>
<td>0.689</td>
</tr>
</tbody>
</table>

### Positive predictive value

<table>
<thead>
<tr>
<th>Image type</th>
<th>Mean</th>
<th>Difference in means (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>0.79</td>
<td>Baseline</td>
<td>-</td>
</tr>
<tr>
<td>OP</td>
<td>0.63</td>
<td>-0.14 (-0.21 to -0.07)</td>
<td>0.000</td>
</tr>
<tr>
<td>FO</td>
<td>0.78</td>
<td>-0.02 (-0.09 to +0.05)</td>
<td>0.516</td>
</tr>
<tr>
<td>WO</td>
<td>0.72</td>
<td>-0.00 (-0.08 to +0.07)</td>
<td>0.936</td>
</tr>
</tbody>
</table>

Table 15 - Results of regression analysis for the focal lesion group alone (n=23). True positives, sensitivity, positive predictive value and confidence are compared across the four image types, using the in phase images as the baseline.
### True positives

<table>
<thead>
<tr>
<th>Image type</th>
<th>Mean</th>
<th>Difference in means (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>4.0</td>
<td>Baseline</td>
<td>-</td>
</tr>
<tr>
<td>OP</td>
<td>2.9</td>
<td>-1.13 (-4.0 to +1.8)</td>
<td>0.449</td>
</tr>
<tr>
<td>FO</td>
<td>5.0</td>
<td>+1.04 (-1.9 to +4.0)</td>
<td>0.483</td>
</tr>
<tr>
<td>WO</td>
<td>4.7</td>
<td>+0.71 (-2.2 to +3.6)</td>
<td>0.634</td>
</tr>
</tbody>
</table>

### Sensitivity

<table>
<thead>
<tr>
<th>Image type</th>
<th>Mean</th>
<th>Difference in means (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>0.22</td>
<td>Baseline</td>
<td>-</td>
</tr>
<tr>
<td>OP</td>
<td>0.15</td>
<td>-0.07 (-0.23 to +0.09)</td>
<td>0.398</td>
</tr>
<tr>
<td>FO</td>
<td>0.29</td>
<td>+0.08 (-0.09 to +0.24)</td>
<td>0.349</td>
</tr>
<tr>
<td>WO</td>
<td>0.25</td>
<td>+0.03 (-0.12 to +0.19)</td>
<td>0.674</td>
</tr>
</tbody>
</table>

### Positive predictive value

<table>
<thead>
<tr>
<th>Image type</th>
<th>Mean</th>
<th>Difference in means (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>0.85</td>
<td>Baseline</td>
<td>-</td>
</tr>
<tr>
<td>OP</td>
<td>0.58</td>
<td>-0.27 (-0.46 to -0.09)</td>
<td>0.004</td>
</tr>
<tr>
<td>FO</td>
<td>0.65</td>
<td>-0.15 (-0.33 to +0.033)</td>
<td>0.113</td>
</tr>
<tr>
<td>WO</td>
<td>0.85</td>
<td>-0.06 (-0.26 to +0.13)</td>
<td>0.522</td>
</tr>
</tbody>
</table>

Table 16 - Results of regression analysis for the diffuse disease group (n=6). True positives, sensitivity, positive predictive value and confidence are compared across the four image types, using the in phase images as the baseline.

In the focal disease group, true positives and sensitivity were significantly higher in the FO group than in the IP group (p=0.008 and 0.037 respectively). There was no significant difference in PPV between IP and FO groups (p=0.516). The OP images performed significantly less well than IP images in terms of PPV (p=0.000).

In the diffuse disease group, there were no significant differences between FO and IP groups in terms of true positives, sensitivity or PPV (p=0.483, p=0.349 and p=0.113 respectively). PPV was again significantly poorer on the OP images than on the IP images (p=0.004).
True positives and sensitivity were higher for the focal disease group (sensitivity on FO images was 0.29 in the diffuse group and 0.46 in the focal group), although these groups were not formally compared.

**Percent Contrast and Contrast-to-noise ratio**

Comparison of percent contrast and CNR between groups is demonstrated in Figure 10-6. Percent contrast was highest in the FO group (the values for each image type were, IP: 8.1, OP: 17, FO: 30 and WO: 4.5), and was significantly higher for FO images than for IP images (p=0.003). There was no significant difference between OP and IP images, or between WO and IP images.

Contrast to noise ratio was also highest in the FO group (the values for each image type were, IP: 3.41, OP: 3.34, FO: 5.57 and WO: 5.04). However, was no significant difference in CNR between groups.

![Figure 10-6 - Comparison of Percent Contrast and CNR between groups. The figures show the results of a post-hoc multiple comparison test from a one-way ANOVA. Estimates of Percent Contrast and CNR are shown as circles; the comparison intervals for each group are shown as solid lines. Percent contrast was significantly higher on FO images than on IP images (p=0.003).](image)

**10.2.4 Discussion**

In this study, lesion counts, true positive counts, sensitivity, positive predictive value and reader confidence were compared across the four Dixon images types. We have shown that FO images are superior to other image types - and in particular IP images - in terms of lesion counts, true positives, sensitivity and confidence.
Furthermore, our data suggest that focal lesions demonstrate greater contrast compared to background marrow on FO images than on IP images, which may account for the superior sensitivity of FO images. The positive predictive values for FO images were similar to those for IP and WO images and higher than those for OP images, suggesting that the increase in sensitivity reflects a true increase in lesion conspicuity rather than a lower reader threshold for lesion identification. The use of FO images offered the greatest advantage for patients with focal lesions, but also provided superior sensitivity in patients with diffuse disease.

The superior performance of FO imaging could occur because myelomatous infiltration of the bone marrow causes a proportionally greater reduction in marrow fat content than in water content. Normal adult bone marrow typically consists of 50-90% fat (224,381,382) and infiltration with myeloma cells decreases fat content (170,271,376); however, the increase in water content may be relatively less because myeloma cells have an increased nuclear to cytoplasmic ratio (383). This suggestion is supported by the observation that focal lesions are more difficult to detect in younger patients with cellular bone marrow imaging (384) or in myeloma patients with a higher bone marrow cell percentage (377).

To our knowledge, this is the first study comparing lesion detection rates on individual Dixon images in patients with MM. A small number of studies have examined lesion contrast in Dixon imaging compared to other sequences (377,385), but none of these have directly examined lesion detection rates by radiologists. This study suggests that the use Dixon imaging improves diagnostic sensitivity and confidence compared to in phase T1-weighted gradient echo imaging alone. We therefore argue that Dixon imaging should be used in preference to T1-weighted imaging alone for anatomical WB-MRI in MM. Furthermore, radiologists should specifically review the FO image type when reading WB-MRI in MM to increase diagnostic yield and improve reporting efficiency.

The accuracy of lesion detection in MM directly impacts on assessment of disease burden and therefore prognosis (371). Walker et al. showed that patients with more than seven focal lesions on WB-MRI had a five year survival of 55%, compared to 73% for those with no focal lesions (371). Moulopoulos et al. similarly showed that radiological assessment of disease burden could be used to separate patients into different survival categories (386). In patients with only a small number of lesions,
poor diagnostic sensitivity could theoretically alter the diagnosis itself – small volume disease could be missed altogether, or patients with a small number of lesions (>1) could be incorrectly diagnosed with solitary plasmacytoma.

A limitation of this study is that our observations are confined to images generated using a single Dixon sequence. It would be preferable to compare sensitivity and positive predictive value across gradient echo (Dixon) and spin echo images including T1-weighted and STIR images, to form a more definitive overall assessment of the optimal sequence. However, this type of study would be difficult to perform in practice since acquiring conventional T1-weighted spin echo images in addition to Dixon images would be extremely time consuming. Furthermore, previous studies suggest that gradient echo imaging offers similar image quality spin echo imaging in MM (378).

The study is also limited by the nature of the scoring system used. In particular, the upper limit of 20 for the lesion count means that we have not captured differences in the number of lesions detected in patients with very high tumour load. However, the clinical importance of these differences is doubtful and current staging systems do not differentiate between patients with more than 20 lesions (365,370). Our scoring system also penalises observers who fail to identify diffuse infiltration, leading to generally low sensitivity scores when compared to the reference standard.

Further work is required to examine the diagnostic utility of different MR sequences to arrive at an optimised protocol for WB-MRI in MM. In particular, it would be useful to determine the extent to which DWI, post-contrast and pre-contrast Dixon imaging each contribute to the overall interpretation of the WB-MRI scan. Careful assessment of the ‘value’ of each sequence is essential if cost-effective, high volume whole body scanning is to be achieved. High-value MRI is becoming an increasingly important goal for the imaging community (387), and studies specifically examining the value of WB-MRI in MM will be essential for widespread clinical implementation.

10.2.5 Conclusion

Fat-only Dixon images offer higher lesion detection rates compared to in-phase images alone in multiple myeloma. We suggest that radiologists should
preferentially review the fat-only images when reading to improve diagnostic accuracy and reporting efficiency.

10.2.6 Acknowledgements


†Denotes equal contribution.
References


20. Gran JT, Skomsvoll JF. The outcome of ankylosing spondylitis: a study of


References


56. Kang KY, Kim IJ, Yoon MA, Hong YS, Park SH, Ju JH, et al. Fat metaplasia on sacroiliac joint magnetic resonance imaging at baseline is associated with spinal radiographic progression in patients with axial spondyloarthritis. PLoS
References


66. Francois, RJ, Neure, L, Sieper, J, Braun J. Immunohistological examination of open sacroiliac biopsies of patients with ankylosing spondylitis: detection of


96. Maksymowych WP. Biomarkers in axial spondyloarthritis. Curr Opin Rheumatol. Lippincott Williams and Wilkins (E-mail: kathiest.clai@apta.org); 2015 Jul;27(4):343–8.


References


117. Rudwaleit M, Jurik AG, Hermann KGA, Landewé R, Van Der Heijde D, Baraliakos X, et al. Defining active sacroiliitis on magnetic resonance imaging


125. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et


134. Baraliakos X, Davis J, Tsuji W, Braun J. Magnetic resonance imaging


idiopathic arthritis: An observational study with patients from the German Methotrexate Registry. Arthritis Care Res. 2012;64(9):1349–56.


References


2012;22(3):276–82.


178. Le Bihan D. Apparent Diffusion Coefficient and Beyond: What Diffusion MR
References


181. Sotak CH. Nuclear magnetic resonance (NMR) measurement of the apparent diffusion coefficient (ADC) of tissue water and its relationship to cell volume changes in pathological states. Neurochem Int. 2004;45(4):569–82.


189. Yeung HN, Kormos DW. Separation of true fat and water images by


209. Janiczek RL, Gambrota G, Sinclair CDJ, Yousry TA, Thornton JS, Golay X,


246. Padhani AR, Hayes C, Landau S, Leach MO. Reproducibility of quantitative


References


273. Shun-Li, O, Nan-Nan, W, Cheng-Lin, S, Jing-Yao, L, Zuo-Wei, L, Shu-Qin G. Investigation of inter-molecular hydrogen bonding in the binary mixture (acetone + water) by concentration dependent Raman study and ab initio


277. Ozcan, M, Seven S. Physical and chemical analysis and fatty acid composition of peanut, peanut oil, and peanut butter from COM and NC-7 cultivars. Grasas y Aceites. 2003;54:12–8.


282. Bollow M, Fischer T, Reishauer H, Backhaus M, Sieper J, Hamm B BJ. Quantitative analyses of sacroiliac biopsies in spondyloarthropathies: T cells and macrophages predominate in early and active sacroilitis- cellularity correlates with the degree of enhancement detected by magnetic resonance


318. Cornell MRI Research Lab. MEDI toolbox.

319. ITK-SNAP.


