

# Development of quantitative MRI as an outcome measure in Charcot- Marie-Tooth disease and inclusion body myositis

---

**Dr Jasper Mark MORROW**

**UCL, Institute of Neurology**

**Primary supervisors:**

**Prof Michael G Hanna, Prof Mary M Reilly**

**Thesis submitted for the degree of Doctor of Philosophy**

**University College London**

**2016**

## Declaration

I, Jasper Mark MORROW 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.

A handwritten signature in black ink, appearing to read 'J. Morrow', written in a cursive style.

## Abstract

Lack of sensitive outcome measures is a major obstacle to clinical trials in many neuromuscular diseases (NMD). Lower limb muscle MRI allows non-invasive visualisation of acute and chronic pathology in NMD. This thesis aims to assess the reliability, validity and responsiveness of quantitative MRI in chronic neuromuscular diseases. A comprehensive quantitative MRI protocol of lower limb muscles was developed including T1, T2, fat fraction and magnetisation transfer ratio (MTR) measurements. The protocol was assessed for reliability and sensitivity to physiological variation in 47 healthy volunteers with 15 rescanned at a two week interval. This protocol was then performed together with detailed clinical assessments and isokinetic/isometric dynamometry in 20 patients with inclusion body myositis (IBM), 20 patients with Charcot-Marie-Tooth disease (CMT) and matched health volunteers twice at a 12 month interval. In the healthy volunteers, the inter-scan and inter-observer reliability was high (ICC 0.62-0.99) despite small observed physiological variation between subjects. Fat fraction, T2 and MTR showed significant correlations with subject age in thigh and calf muscles and with subject weight in thigh muscles whereas gender did not influence quantitative parameters. Cross-sectional analysis showed strong correlations with both muscle strength and clinical severity measures demonstrating validity of MRI measurements as outcome measures. Longitudinal assessment demonstrated excellent sensitivity to change of MRI measures; in particular muscle fat fraction quantification exceeded that of myometry and clinical measurements with standardised response mean (SRM) over 12 months of 1.1 in IBM and 0.8 in CMT indicating a high level of responsiveness. Annual change in fat fraction could be predicted based on baseline MRI measurements, providing the opportunity to improve SRM further. This thesis demonstrates the reliability, validity and responsiveness of quantitative MRI as an outcome measure providing a comprehensive practical protocol for clinical trials in NMD.

## Acknowledgements

Collaboration between neurologists, radiologists and clinical scientists is vital for neuromuscular MRI research, and this PhD would not have been possible without the valuable input of many colleagues. Whilst it is noted in the text where specific aspects have been performed by others, I would like to take the opportunity to acknowledge their wider input into this PhD.

Dr Christopher Sinclair provided MRI physics expertise to the project, with specific involvement in study design, MRI protocol implementation and MRI map processing. Above and beyond this, he provided support and advice whenever it was needed, without which this PhD would not have been possible. The visiting radiologist, Dr Arne Fischmann, performed the ROI definition on quantitative sequences and Mercuri grading of qualitative images, which underlies all of the MRI results presented in this thesis. This involvement continued even after his return to Switzerland, for which I am extremely grateful. Dr Adrian Miller and Dr Pedro Machado were involved in study design and obtaining ethical approval for the study prior to my arrival, which allowed the data collection to get underway promptly.

My joint primary supervisors, Professor Mary Reilly and Professor Michael Hanna provided complementary support throughout the PhD. It was a privilege to work under their supervision. I'd like to thank Dr John Thornton for supervision of the MRI physics aspects, and Professor Tarek Yousry for his supervision and support of the radiological aspects of the project. I am grateful for the input of all my supervisors in the preparation of the published manuscripts and thesis.

I would like to thank the study participants, who volunteered their time without reimbursement to take part in this study. I would like to thank my father for meticulous proof editing of the entire thesis. And finally I would like to thank my wife for all her support.

# Table of contents

Declaration	2
Abstract	3
Acknowledgements	4
Table of contents	5
List of tables	14
List of figures	16
List of equations	18
List of abbreviations	19
1 Background .....	21
1.1 Introduction	21
1.2 Neuromuscular diseases: the need for an outcome measure	21
1.2.1 General principles of outcome measures.....	22
1.2.2 Challenges for outcome measures in NMD.....	25
1.2.3 Possible outcome measures in neuromuscular diseases.....	26
1.2.4 Case study of use of an outcome measure in a clinical trial in NMD .....	28
1.2.5 MRI as an outcome measure.....	29
1.3 Principles of qualitative neuromuscular MRI	30
1.3.1 MRI as applied to muscle and nerve.....	30
1.3.2 MRI in inherited muscle diseases .....	32
1.3.3 MRI in inflammatory muscle diseases.....	33
1.3.4 Muscle MRI in inherited and acquired neuropathies .....	33
1.4 Quantitative neuromuscular MRI	34
1.4.1 T1 and T2 mapping .....	34
1.4.2 Fat quantification .....	35
1.4.3 Magnetisation transfer imaging.....	36

1.5	Aims of thesis	37
2	Methods.....	38
2.1	Background	38
2.1.1	Development of novel outcome measures .....	38
2.1.2	Disease group selection .....	39
2.1.2.1	Inclusion body myositis	40
2.1.2.2	Charcot-Marie-Tooth disease 1A	40
2.1.3	Use of healthy controls .....	40
2.1.4	Sample size.....	41
2.1.5	Ethical framework.....	42
2.2	Study design	42
2.2.1	Subject recruitment.....	43
2.3	Clinical assessments	43
2.3.1	Disease specific scales.....	44
2.3.2	Myometry.....	44
2.3.2.1	Myometry analysis	45
2.3.3	MRC strength grading.....	45
2.3.4	Other assessments.....	46
2.4	MRI protocol	46
2.4.1	Hardware.....	46
2.4.2	Subject positioning .....	46
2.4.3	Anatomical coverage .....	47
2.4.4	MRI sequences.....	47
2.4.4.1	Fat fraction measurement	47
2.4.4.2	T <sub>1</sub> -Relaxometry	48
2.4.4.3	T <sub>2</sub> -Relaxometry	48

2.4.4.4	B <sub>1</sub> Mapping	49
2.4.4.5	Magnetization transfer ratio	49
2.4.4.6	T1 weighted imaging	50
2.4.4.7	Short-tau-inversion-recovery imaging	50
2.5	Qualitative image analysis	50
2.6	Quantitative analysis	50
2.6.1	Slice selection .....	50
2.6.2	Region of interest definition .....	51
2.6.3	Application of region of interest to maps .....	52
2.7	Data analysis	52
2.7.1	Data checking.....	53
2.7.2	Statistical analysis .....	53
3	Quantitative MRI in healthy volunteers.....	55
3.1	Background	55
3.1.1	Ideal outcome measure characteristics in healthy volunteers .....	56
3.2	Quantitative MRI variability and reliability in healthy volunteers	57
3.2.1	Published values and variation between muscles .....	57
3.2.1.1	T2 time	58
3.2.1.2	T1 time	61
3.2.1.3	Fat fraction	61
3.2.1.4	MTR	66
3.2.2	Physiological variation .....	66
3.2.2.1	Effect of exercise and training	67
3.2.2.2	Effect of diet	67
3.2.3	Inter-subject variation .....	68
3.2.3.1	Effect of age on quantitative MRI parameters	68
3.2.3.2	Effect of gender on quantitative MRI parameters	69

3.2.3.3	Effect of body habitus on MRI parameters	70
3.2.4	Reliability .....	70
3.2.5	Summary .....	72
3.3	Healthy volunteer phase aims	72
3.4	Results	73
3.4.1	Subjects and scan schedule .....	73
3.4.2	General appearance .....	73
3.4.3	Individual muscle values .....	75
3.4.4	Scan-rescan and inter-observer reliability .....	77
3.4.4.1	Muscle area	77
3.4.5	Dependence upon age, gender and weight .....	86
3.4.6	Relationship between quantitative parameters .....	89
3.4.7	Qualitative imaging .....	90
3.5	Discussion	93
3.5.1	Intramuscular fat accumulation .....	93
3.5.1.1	Dixon fat fraction	93
3.5.1.2	T1 measurements	96
3.5.1.3	Qualitative assessment of T1-weighted images	96
3.5.2	Correlation between quantitative MRI measurements .....	99
3.5.3	Water sensitive measures .....	100
3.5.3.1	T2 measurement	100
3.5.3.2	MTR measurement	102
3.5.3.3	Qualitative STIR imaging	103
3.5.4	Muscle size measurements .....	103
3.5.4.1	Cross-sectional muscle area	103
3.5.5	Reproducibility .....	104
3.5.5.1	Reproducibility of measures of intramuscular fat accumulation	105

3.5.5.2	Reproducibility of water sensitive measures	106
3.5.5.3	Inter-observer reliability	106
3.5.5.4	Methods to optimise reliability	107
3.5.6	Feasibility/Study limitations.....	110
3.6	Conclusions	111
4	Cross-sectional quantitative MRI in neuromuscular disease .....	113
4.1	Introduction	113
4.1.1	Validity.....	113
4.2	Background literature	115
4.2.1	Qualitative MRI studies in CMT1A/IBM.....	115
4.2.1.1	Published reports in CMT1A	115
4.2.1.2	Published qualitative MRI reports in IBM	116
4.2.2	Studies of quantitative MRI in neuromuscular diseases .....	117
4.2.2.1	Quantitative MRI measures of chronic muscle pathology	117
4.2.2.2	Quantitative measures of acute muscle pathology in neuromuscular diseases	119
4.2.3	Correlation between MRI parameters .....	120
4.2.3.1	Quantitative vs qualitative	120
4.2.3.2	Quantitative vs quantitative	120
4.2.4	Correlation with clinical parameters .....	120
4.2.4.1	Strength	120
4.2.4.2	Functional measures	121
4.2.4.3	Neurophysiology	122
4.2.4.4	Pathology	122
4.2.5	Summary.....	123
4.3	Cross-sectional data aims	123
4.4	Results	124
4.4.1	Subjects .....	124

4.4.2	MRI data.....	124
4.4.3	Measures of chronic fatty atrophy.....	125
4.4.3.1	Qualitative T1w sequences	126
4.4.3.2	Overall quantitative MRI results	127
4.4.3.3	Distribution of MRI abnormalities	128
4.4.3.4	MRI correlations	134
4.4.4	Probing water distribution abnormalities .....	135
4.4.4.1	Qualitative STIR sequences	136
4.4.4.2	Early water abnormalities in muscles without significant intramuscular fat accumulation	137
4.4.4.3	Assessment of T2 values relative to FF levels	139
4.4.5	Clinical data.....	141
4.4.5.1	Myometry	141
4.4.6	MRI - clinical correlations.....	143
4.4.6.1	Correlations between overall MRI and clinical measures	144
4.4.6.2	Correlation between MRI measures and myometry	146
4.4.6.3	Assessment of muscle quality	149
4.5	Discussion	151
4.5.1	Validity of MRI measures of fatty atrophy.....	151
4.5.1.1	Construct validity	151
4.5.1.2	Criterion validity	153
4.5.1.3	Face validity	154
4.5.1.4	Content validity	154
4.5.1.5	Validity conclusions	155
4.5.2	Measures of abnormal muscle water distribution .....	155
4.5.3	Insights into CMT1A .....	156
4.5.4	Insights into IBM.....	159
4.5.5	Challenges and study limitations .....	161

4.5.6	Future application.....	162
4.6	Conclusions	163
5	Longitudinal quantitative MRI in neuromuscular disease .....	164
5.1	Introduction	164
5.2	Background literature	165
5.2.1	Longitudinal qualitative MRI studies in CMT/IBM.....	165
5.2.2	Natural history quantitative MRI studies of muscle size .....	165
5.2.3	Interventional studies including muscle size measurements .....	165
5.2.4	Natural history studies of quantitative tissue parameters .....	167
5.2.5	Interventional studies including quantitative muscle tissue parameters	169
5.2.6	Summary.....	170
5.3	Longitudinal study aims	171
5.4	Results	172
5.4.1	Study subjects.....	172
5.4.2	Overall longitudinal data .....	172
5.4.3	Longitudinal assessment of fat atrophy.....	174
5.4.3.1	Qualitative imaging	175
5.4.3.2	Quantitative imaging	175
5.4.4	Longitudinal clinical data.....	177
5.4.4.1	CMT1A	177
5.4.4.2	IBM	177
5.4.5	Longitudinal MRI - clinical correlation .....	177
5.4.6	Predictors of change.....	178
5.4.6.1	Baseline T2 as predictor of change	179
5.4.7	Standardised Response Mean.....	180
5.5	Discussion	183

5.5.1	Outcome measure responsiveness.....	183
5.5.1.1	Responsiveness of outcome measures in CMT1A	183
5.5.1.2	Responsiveness of outcome measures in IBM	185
5.5.2	Baseline predictors of change.....	185
5.5.3	Longitudinal insights into CMT1A .....	185
5.5.4	Longitudinal insights into IBM .....	187
5.5.5	Longitudinal study limitations .....	189
5.6	Conclusions	189
6	Conclusions.....	190
6.1	Specific conclusions and future directions	192
6.1.1	Study design.....	192
6.1.2	MRI protocol .....	193
6.1.2.1	MRI hardware	193
6.1.2.2	Scan scheduling and subject positioning	194
6.1.2.3	Anatomical coverage and slice positioning	195
6.1.2.4	MRI sequences	197
6.1.2.5	Novel methods which might be considered for future studies	198
6.1.3	Analysis methods .....	199
6.1.3.1	General considerations	199
6.1.3.2	Z-axis considerations	200
6.1.3.3	X-Y plane considerations	204
6.1.3.4	Data checking	205
6.1.4	Outcome measure selection and trial design .....	205
6.1.4.1	CMT1A	205
6.1.4.2	IBM	207
6.1.5	Application to clinical trials of other diseases .....	208
6.1.5.1	Quantitative MRI protocol in hypokalaemic periodic paralysis	209
6.1.5.2	Quantitative MRI protocol in Hereditary Sensory Neuropathy	209

6.1.5.3	Quantitative muscle MRI protocol in other neuromuscular diseases	
	210	
6.2	Final conclusions	210
7	Publications related to this PhD .....	212
8	References .....	214

## List of tables

Table 1-1: Sample size by SRM and efficacy .....	25
Table 1-2: Neuromuscular MRI Tissue Characteristics .....	31
Table 1-3: Mercuri Scale.....	33
Table 2-1: Myometry protocol .....	45
Table 3-1: Reported T2 in skeletal muscle of healthy volunteers .....	60
Table 3-2: T2 times in different calf muscles.....	61
Table 3-3: Reported T1 values in skeletal muscle of healthy volunteers .....	61
Table 3-4: Reported fat fraction in skeletal muscle of healthy volunteers .....	65
Table 3-5: Reported MTR in the skeletal muscle of healthy volunteers.....	66
Table 3-6: Overall reliability table: inter-scan and inter-observer.....	79
Table 3-7: Mean quantitative parameters in 15 healthy volunteers and test-retest reliability .....	80
Table 3-8: Inter-observer reliability of MRI measurements.....	81
Table 3-9: Inter-scan paired two-tailed t-tests for each modality for each ROI .....	82
Table 3-10: Inter-observer paired two-tailed t-tests for each modality for each ROI ....	83
Table 3-11: Limits of agreement (LOA) for each quantitative parameter for each ROI individually.....	85
Table 3-12: Multivariate regression analysis of muscle MRI measures with subject demographics.....	87
Table 3-13: Multivariate regression of individual muscle MRI measures with demographic factors .....	88
Table 3-14: Pearson correlation coefficients between quantitative parameters .....	89
Table 3-15: Frequency of Mercuri grades in different muscle groups.....	90
Table 3-16: Demographic and MRI associations of right medial gastrocnemius STIR hyperintensity .....	92
Table 3-17: Advantages of different ROI types .....	95
Table 3-18: Conversion of Mercuri Scales .....	98

Table 4-1: Subject demographics .....	124
Table 4-2: MRI data summary results at thigh and calf level .....	125
Table 4-3: Distribution of qualitative Mercuri grading by subject group.....	127
Table 4-4: Cross-sectional area by muscle in patients and matched controls .....	128
Table 4-5: Muscle-specific MRI parameters in CMT1A and matched controls.....	129
Table 4-6: Muscle-specific MRI parameters in IBM and matched controls .....	130
Table 4-7: Correlation of quantitative MRI parameters .....	135
Table 4-8: Frequency of STIR grades in thigh and calf muscles .....	136
Table 4-9: Quantitative parameters in muscles without significant IFA.....	138
Table 4-10: Quantitative parameters in muscles without significant IFA grouped by STIR .....	138
Table 4-11: Linear regression T2 and MTR in muscles fat fraction in the healthy control range .....	139
Table 4-12: Clinical data summary.....	141
Table 4-13: Baseline myometry statistics by group .....	141
Table 4-14: Baseline isokinetic and isometric myometry by group .....	143
Table 4-15: Correlation of muscle strength with MRI parameters in each subject group .....	147
Table 4-16: Correlates of tibialis anterior muscle quality .....	150
Table 5-1: Overall longitudinal data .....	173
Table 5-2: Annual change in fat fraction within muscles varies with baseline characteristics.....	178
Table 5-3: Change in FF grouped by "T2-excess" at baseline.....	180
Table 5-4: Standardised Response Means in IBM .....	181
Table 5-5: Standardised Response Means in CMT1A .....	182
Table 5-6: Comparison of outcome measure responsiveness with prior CMT1A studies .....	184

## List of figures

Figure 1-1: Lower Limb Anatomy .....	32
Figure 1-2: Three-point Dixon .....	36
Figure 2-1: Outcome Measure Development Pipeline .....	39
Figure 2-2: Flow chart of subject assessments and drop out for phases 2-3 .....	42
Figure 2-3: Exclusion criteria for all subjects .....	43
Figure 2-4: Analysis of Upgrade effects on T2 .....	49
Figure 2-5: Region of interest definition .....	51
Figure 3-1: Sample images from a single healthy volunteer (24 year old male) .....	74
Figure 3-2: Individual muscle ROI values at thigh and calf levels for 47 healthy subjects .....	76
Figure 3-3: Bland-Altman plots for inter-scan and inter-observer analyses.....	84
Figure 3-4: Relationship between MTR and age .....	86
Figure 3-5: Thigh area vs age in males and females.....	86
Figure 3-6: Scatter-plot showing correlation between fat fraction and T2 time in the calf .....	89
Figure 3-7: Box-plot of whole muscle Dixon versus Mercuri grade for thigh and calf ...	91
Figure 3-8: STIR hyperintensity in medial gastrocnemius .....	91
Figure 3-9: Frequency of STIR hyperintensity in calf muscles of volunteers.....	91
Figure 3-10: Box-plot of T2 times in calf muscles grouped by qualitative STIR classification .....	92
Figure 3-11: Original Mercuri grades.....	97
Figure 4-1: Sample images.....	124
Figure 4-2: Qualitative imaging and Regions of Interest.....	126
Figure 4-3: Distribution of intramuscular fat accumulation in thigh of IBM patients ....	132
Figure 4-4: Qualitative and quantitative pattern of IFA in IBM .....	132
Figure 4-5: Distribution of intramuscular fat accumulation in calf muscles of CMT1A	133
Figure 4-6: Correlation between FF and Mercuri grade.....	134

Figure 4-7: Distribution of STIR abnormalities in IBM patients .....	137
Figure 4-8: T2 vs Fat Fraction in thigh muscles of IBM patients grouped by qualitative STIR rating .....	140
Figure 4-9: Thigh FF range vs T2 in IBM patient grouped by STIR positivity.....	140
Figure 4-10: Correlation of fat fraction with disease duration in IBM and CMT1A patients.....	144
Figure 4-11: Correlation between IBM-FRS (lower limb) and thigh FF in IBM patients .....	145
Figure 4-12: Relationship between age and calf FF in CMT1A patients .....	145
Figure 4-13: Correlation of calf fat fraction with severity score (CMTES) in CMT1A patients.....	145
Figure 4-14: Correlation matrix of MRI FF and IBM-FRS domains in IBM patients....	146
Figure 4-15: Correlation of MRI measured remaining muscle area with isometric muscle strength .....	148
Figure 4-16: Muscle quality of tibialis anterior in the three subject groups.....	149
Figure 4-17: Regression analyses of tibialis anterior muscle quality for IBM patients	150
Figure 4-18: Distribution of IFA in IBM in the literature (left) and this thesis (right) ....	160
Figure 5-1: Longitudinal fat fraction images .....	174
Figure 5-2: Box and whisker plot of longitudinal change in fat fraction .....	175
Figure 5-3: Correlation of change in strength and remaining muscle area on MRI ....	178
Figure 5-4: Longitudinal change in thigh muscles in IBM patients grouped by baseline FF and STIR .....	179
Figure 5-5: Correlation between T2 and FF at baseline to estimate "baseline T2 excess" .....	180
Figure 5-6: A model of disease progression in CMT1A .....	186
Figure 5-7: A model of disease progression in IBM.....	188
Figure 6-1: Effect of slice selection on test-retest reliability .....	202
Figure 6-2: Methods to Improve test-retest reliability.....	203

## List of equations

Equation 1-1: General Reliability Formula.....	23
Equation 1-2: Standardised Response Mean Formula.....	24
Equation 1-3: Lehr's Formula.....	24
Equation 3-1: Fat fraction.....	62
Equation 4-1: Muscle quality formula .....	149

## List of abbreviations

ALS FRS	ALS Functional Rating Scale
AM	Adductor magnus
BF	Biceps femoris
CMT	Charcot-Marie-Tooth disease
CMT1A	CMT type 1A
CMTES	CMT Examination Score
CMTNS	CMT Neuropathy Score
CPMG	Carr-Purcell-Meiboom-Gill
CSA	Cross-sectional area
CT	Computerised tomography
DMD	Duchenne muscular dystrophy
E	Expected efficacy of treatment
EMCL	Extramyocellular lipid
ENMC	European Neuromuscular Centre
ERT	Enzyme replacement therapy
FDA	Federal Drug Agency
FF	Fat fraction
G	Gracilis
GSK	Glaxo-Smith-Kline
IBM	Inclusion body myositis
IBM FRS	IBM Functional Rating Scale
ICC	Intra-class correlation coefficient
IDEAL	Iterative decomposition of water and fat with echo asymmetry and least-squares estimation
IFA	Intramuscular fat accumulation
LG	Lateral head of gastrocnemius
MG	Medial head of gastrocnemius
MRC	Medical Research Council
MRI	Magnetic Resonance Imaging
MRS	Magnetic Resonance Spectroscopy
MTR	Magnetic Transfer Ratio
Pearson ICC	Pearson Inter-class correlation coefficient
RF	Rectus femoris
RMA	Remaining muscle area
ROI	Region of interest
Sa	Sartorius

SM	Semimembranosus
So	Soleus
SRM	Standardised Response Mean
STIR	Short-tau Inversion Recovery sequence
STIR	Semitendinosus
SV	Single voxel
T1	Longitudinal T1 recovery time
T1w	T1-weighted sequence
T2	Transverse T2 relaxation time
TA	Tibialis anterior
TE	Echo time
TR	Repetition time
UCL	University College London
VI	Vastus intermedius
VL	Vastus lateralis
VM	Vastus medialis

# 1 Background

## 1.1 *Introduction*

Advances in knowledge of the pathogenesis of neuromuscular diseases (NMD) have led to the identification of potential therapies which require rigorous assessment in clinical trials. Unfortunately, current outcome measures have limitations in reliability, validity or responsiveness which hamper translation of these advances to therapies for patients. In other areas of neurology, most notably multiple sclerosis (Sahraian and Eshaghi, 2010), magnetic resonance imaging (MRI) is well established as a provider of sensitive surrogate outcome measures. MRI in neuromuscular diseases is already established as a means to visualise both acute and chronic muscle pathology, allowing qualitative assessment to aid in diagnosis (Wattjes et al., 2010). MRI therefore provides an excellent potential source of outcome measures in neuromuscular diseases, but requires the demonstration of reliability, validity and responsiveness prior to utilisation in clinical trials.

This PhD thesis provides a comprehensive, systematic assessment of the reliability, validity and responsiveness of quantitative MRI measurements in neuromuscular diseases. Chapter 2 provides the methods for all of the work presented. The fundamental properties of quantitative MRI of lower limb muscles, including measurement reliability, were assessed in healthy volunteers (chapter 3). The performance in patients with chronic neuromuscular diseases, including validity as an outcome measure, was assessed in a cross-sectional study in patients with inclusion body myositis (IBM) and Charcot-Marie-Tooth disease (CMT), chosen to be representative of chronic neuromuscular diseases (chapter 4). Finally, and perhaps most importantly the responsiveness or sensitivity to change of quantitative MRI outcome measures was assessed in these patients longitudinally (chapter 5). Published literature relevant to each of these areas will be included in each section, whilst the remainder of this first section will provide necessary background to outcome measures and neuromuscular MRI more generally.

## 1.2 *Neuromuscular diseases: the need for an outcome measure*

Knowledge in inherited neuromuscular diseases in the past 25 years has expanded exponentially following the scientific breakthroughs which allowed the identification of the causative genetic defects such as the mutations in dystrophin gene which causes Duchenne muscular dystrophy (DMD) in 1986 (Hoffman et al., 1987), mutations in the sodium channel gene which cause hyperkalaemic periodic paralysis in 1990 (Fontaine

et al., 1990) and the 1.4Mb duplication on chromosome 17 that is responsible for CMT type 1A (CMT1A) in 1991 (Lupski et al., 1991; Raeymaekers et al., 1991). Over the same time period there have been similar advances in knowledge of the pathogenesis of inflammatory neuropathies (Lunn and Willison, 2009) and inflammatory myopathies (Dalakas, 2010).

Together with these advances have come potential therapies for neuromuscular diseases. Some therapies have proven efficacy mostly for the inflammatory conditions, with the notable exception of IBM for which there is no known effective treatment (Dalakas, 2010). In the inherited neuromuscular conditions there are a few therapies which improve symptoms, in particular for the skeletal muscle channelopathies (Raja Rayan and Hanna, 2010); however with the exceptions of alglucosidase alfa for Pompe disease (Kishnani et al., 2007) and corticosteroids in DMD (Manzur et al., 2008) there are no medications which have been shown to alter disease progression. Clinical trials in other neuromuscular diseases have had a negative outcome despite encouraging experimental data such as in recent large randomised trials of vitamin C in CMT1A (Pareyson et al., 2011a). There are two possible reasons for a negative trial result: either the medication is not effective or the trial was not adequately designed to assess effectiveness. In particular, a major issue in neuromuscular diseases is the difficulty in identifying suitable outcome measures, which may in part explain the difficulty in translating the scientific advances into patient benefit.

### **1.2.1 General principles of outcome measures**

Whilst a health measurement instrument can be judged in terms of two basic principles: validity and reliability (Streiner and Norman, 2008), when intended as an outcome measure an additional vital characteristic is responsiveness, defined as the sensitivity of the outcome measure to detect change (Kirshner and Guyatt, 1985). This in part relates to measurement reliability, but also depends on the likely magnitude of change in the measurement over the duration of the clinical trial (Streiner and Norman, 2008).

The ideal outcome measure in terms of validity is one which unequivocally reflects tangible benefit to the patient (De Gruttola et al., 1997), which can be referred to as a “true end-point”. The alternative is a “surrogate end-point”, which can be defined simply as one used in lieu of the true end-point (Wittes et al., 1989). However, for a surrogate end-point to be valid it needs to be shown to correlate with the true-end point. Furthermore, a convincing surrogate should have biological relevance and face validity. Statistical relationship alone is insufficient (Wittes et al., 1989). The primary motivation for surrogate end-points is to reduce trial duration so that trials are feasible (Prentice, 1989). However caution is needed when interpreting trials based on

surrogate outcome measures, as there are numerous examples in medical research where medications with demonstrated efficacy on a surrogate end-point subsequently fail when assessed against a true end-point (Fleming and DeMets, 1996).

As a general rule end-points with higher validity are less sensitive to change, so choosing the most appropriate end-point for a study depends on balancing these two factors. More sensitive surrogate end-points are often used in early-phase clinical trials (Phase I and II) to screen for efficacy, whereas regulatory bodies usually require true end-points to be used in phase III clinical studies prior to drug approval. This is not absolute; the FDA for example will allow approval for a new drug under some circumstances if established that the drug has “an effect on a surrogate endpoint that is reasonably likely, based on epidemiologic, therapeutic, pathophysiologic, or other evidence, to predict clinical benefit or on the basis of an effect on a clinical endpoint other than survival or irreversible morbidity” (Code of Federal Regulations, 2010). Even in these circumstances post-marketing studies must show benefits in a true end-point or else drug approval can be withdrawn.

Reliability of an outcome measure as traditionally expressed with an intra-class correlation coefficient (ICC) is not fixed but dependent on study design; in particular reliability studies with a more heterogeneous population will have greater ICCs with an otherwise identical trial design (Streiner and Norman, 2008). This is as reliability is defined as the ratio of inter-subject variation to inter-subject variation plus measurement error so will appear higher (closer to 1 - nominally perfect reliability) if inter-subject variation is greater.

This has led some health researchers to criticise the classical statistical method of quantifying reliability. In particular Bland and Altman (Bland and Altman, 1986) developed a method for use in health research which defines reliability according to the limits of agreement independently of the overall inter-subject variation, which they felt was more meaningful in a medical setting. The results however can be directly obtained from the ICC, so are in reality a different way of expressing the same statistical method (Streiner and Norman, 2008).

$$\text{Reliability} = \frac{\text{Intersubject Variation}}{\text{Intersubject variation} + \text{Random Variation}}$$

*Equation 1-1: General Reliability Formula*

Depending on the outcome measure and study design chosen, reliability will be a combination of test-retest reliability, inter-observer reliability if more than one observer and inter-site reliability if more than one site. The reason that reliability is so important when a measurement instrument is to be used as an outcome measure is that it directly contributes to variation in apparent change seen between subjects. This

variation has a significant negative effect on the responsiveness of an outcome measure. Put simply, a large measurement error creates “noise” which can make it impossible to detect small amounts of change.

Whilst responsiveness of an outcome measure to detect change can be expressed in many ways (Streiner and Norman, 2008), perhaps the most appropriate for chronic neuromuscular diseases is the standardised response mean (SRM) originally described by Liang and colleagues (Liang et al., 1990). This can be derived from data from a natural history study and is a ratio calculated as the mean change over the duration of the study ( $\delta$ ) divided by the standard deviation of the change ( $\sigma$ ) (Equation 1-2: Standardised Response Mean Formula

). Whether the SRM is negative or positive is unimportant, it is the magnitude which describes responsiveness. An absolute value of less than 0.2 is considered to have minimal responsiveness, between 0.2 and 0.5 small responsiveness, 0.5-0.8 moderate responsiveness and >0.8 large responsiveness.

$$\text{SRM} = \frac{\delta}{\sigma}$$

*Equation 1-2: Standardised Response Mean Formula*

The SRM is a key factor in the power of a study as it determines (along with the efficacy of the intervention), the sample size required for a clinical study. This can be seen with Lehr’s formula(Lehr, 1992) (Equation 1-3: Lehr’s Formula

) which estimates sample size required for a study:

$$N = \frac{16}{(\text{SRM} \times E)^2}$$

N = number required in each group for a study of 80% power with significance level  $p < 0.05$

E = expected efficacy of treatment compared with natural history – e.g. 0.3 = 30% reduced progression, 1.0 = halts progression, 1.5 = halts progression and improves at 50% of rate of progression (Petrie and Sabin, 2005)

*Equation 1-3: Lehr’s Formula*

Table 1-1 lists sample sizes depending on SRM and treatment efficacy and demonstrates how important maximising SRM is in limiting the sample size needed for a clinical trial. It is worth noting at this point that the standard deviation of the change in the outcome measure is a combination of the variation due to true variation in rate of progression between patients, and the apparent variation due to measurement error –

determined by the reliability of the measurement instrument. In practical terms therefore the responsiveness of an outcome measure can be maximised by choosing one which changes as much as possible over the length of the trial and for this change to be consistent between subjects by limiting both true and erroneous inter-subject variation.

*Table 1-1: Sample size by SRM and efficacy*

SRM	Treatment effect on chosen outcome measure (E)			
	30% reduced progression (0.3)	60% reduced progression (0.6)	Halts progression (1.0)	Improves 50% (1.5)
0	Infinity	Infinity	Infinity	Infinity
0.1	17778	4444	1600	711
0.2	4444	1111	400	178
0.4	1111	278	100	44
0.6	494	123	44	20
0.8	278	69	25	11
1	178	44	16	7
2	44	11	4	2

*Sample size required in each group for a clinical trial of power 0.8,  $p < 0.05$ , in a progressive disease as determined by standardised response mean (SRM) of outcome measure and efficacy of treatment.*

## 1.2.2 Challenges for outcome measures in NMD

Many challenges confront a potential outcome measure for neuromuscular diseases.

### **Progression is slow**

Most neuromuscular diseases are non-fatal, with a few notable exceptions including Guillain-Barré syndrome and motor neuron disease. This means patient survival cannot be used as a true outcome measure as defined by the FDA, but an outcome measure must demonstrate prevention of irreversible morbidity (Code of Federal Regulations, 2010). However, the slow progression of neuromuscular diseases necessitates long trial duration. This increases expense and causes delays to treatment licensing for patient use.

### **Treatments of hereditary or degenerative disorders are likely to primarily halt progression rather than cause improvement**

Axonal loss in peripheral nerves and fatty atrophy in muscles are unlikely to be easily reversed by treatments, which are more likely to slow or halt progression. This limits the magnitude of difference which may be seen between treatment and placebo arms in a clinical study. This means an even more reliable outcome measurement is needed for a study of an inherited or degenerative disease than in an inflammatory myopathy

or neuropathy where the treatment group may improve while the placebo group worsens (right column Table 1-1). A potential advantage however is that the natural history of progression is relatively constant between patients and spontaneous improvement unlikely, which lessens inter-subject variation in longitudinal change. In CIDP or multiple sclerosis, fluctuation of disability levels due to the disease process itself can mask the beneficial effect of an intervention. To maximise this advantage, measurement error must be minimised otherwise apparent improvement may be seen due to measurement error (poor test-retest reliability) rather than a true improvement, which similarly masks beneficial effects.

### **Neuromuscular conditions are rare**

Neuromuscular diseases are relatively rare compared with some other disease groups, such as cardiovascular disease. This limits recruitment and means a multi-centre or even a multi-national trial design may be needed. This means both that an outcome measure needs to be evaluated in this trial design, and furthermore the addition of inter-site variability will reduce outcome measure reliability. Rarity also makes disease-specific scale validation more difficult. The alternative is to use a generic scale, however these are generally less sensitive to change in specific patient populations (Streiner and Norman, 2008).

### **1.2.3 Possible outcome measures in neuromuscular diseases**

This section will briefly review previously used and potential outcome measures in NMD in the context of the difficulties noted above. Ratings in brackets represent the subjective ratings of the thesis author on a scale ranging A-E to indicate the relative strengths and weaknesses of current outcome measures.

#### **Patient survival (Validity = A, Reliability = A, Responsiveness = E)**

This is only practical in a small number of diseases such as motor neuron disease or severe Guillain-Barré syndrome. It could potentially be applied to a study in DMD, but only to those with late stage disease where disease specific interventions are less likely to be effective. Furthermore, using death as an outcome measure requires a large sample size which would not be practical in most neuromuscular diseases

#### **Disability (Validity A, Reliability B, Responsiveness D)**

Irreversible morbidity is considered a true outcome measure by the FDA (Code of Federal Regulations, 2011) so a direct measure of disability scores highly for validity as an outcome measure in neuromuscular disease. However, disability measurement isn't as straightforward as might first appear. Even something as simple as wheelchair use is not dichotomous but a spectrum from occasional use to complete dependence.

Even the “date of first wheelchair use” is not solely determined by patient disability, but also by social and economic factors such as cost and availability of a wheelchair, and the patient’s willingness to use a wheelchair. If rather than a dichotomous event, one instead decides to use a disability scale, there is the question of whether to use a disease-specific scale which may be more likely to detect change in the patients in question, or a generic scale which is generalisable across different conditions – known as the ‘fidelity versus bandwidth’ issue (Streiner and Norman, 2008), with advantages and disadvantages for each approach. Perhaps most importantly regardless of which kind of measure is used, with a gradually progressive disease disability measures will change in a minority of patients over the likely time-course of a clinical trial so responsiveness will be poor.

### **Disease-specific composite scales (Validity C, Reliability B, Responsiveness D)**

Disease-specific scales are attractive as they can combine many different disease measurements in a single score. Validated measures exist for a number of chronic neuromuscular conditions such as the CMT Neuropathy Score (Shy et al., 2005), which has been used as an outcome measure in recent vitamin C trials. However, they may contain aspects such as sensory examination which have poor inter-observer reliability and neurophysiology components which have less face validity. Both individual components and the overall score have been shown to be relatively insensitive to change as demonstrated in the case study which follows (Pareyson et al., 2011a).

### **Patient reported outcome measures (Validity B, Reliability C, Responsiveness C)**

These are defined as any outcome measure which the study participant self-administers such as quality of life questionnaires or other self-rating scales (Streiner and Norman, 2008). These certainly have an important place as secondary outcome measures in a clinical trial, but are affected by many non-disease specific factors and are unlikely to be sufficiently sensitive to show a significant difference in a chronic neuromuscular disease trial.

### **Muscle strength (Validity C, Reliability C, Responsiveness C)**

Muscle strength is perhaps the most logical way to directly measure disease severity in chronic neuromuscular disorders, with the exception of pure or mainly sensory neuropathies. However, strength measurement is not as straightforward as might be imagined. Many studies have used manual muscle testing such as the Medical Research Council (MRC) scale of a varied combination of muscle groups as a primary or secondary outcome measure (Peng et al., 2000; Walter et al., 2000), but it has been shown that MRC scores have relatively poor inter-observer reliability, especially between mid-range scores (Vanhoutte et al., 2011). Hand-held dynamometry has also

been used in a number of studies but test-retest and inter-rater reliability can be poor (Solari et al., 2008). Fixed myometry may be more reliable but is time consuming and has been used in fewer neuromuscular trials (Dalakas et al., 1997). Regardless of the method, accurate measurement is dependent on maximal patient effort which may vary between visits and affect results considerably.

### **Investigations (Validity C-D, Reliability B-C, Responsiveness B)**

Whilst creatine kinase level, muscle biopsy, skin biopsy and neurophysiology have a critical and irreplaceable role in diagnosis, their use as surrogate outcome measures remains to be established. Except in early phase studies, laboratory investigations have generally been used as secondary outcome measures in neuromuscular trials because they lack face validity. Muscle or skin biopsy is invasive and samples only a small area tissue. Neurophysiology can vary significantly between observers and longitudinal studies suggest it lacks sensitivity to change (Verhamme et al., 2009).

#### **1.2.4 Case study of use of an outcome measure in a clinical trial in NMD**

The Italian-British trial of ascorbic acid in CMT1A can be considered to exemplify the difficulties facing outcome measures used in clinical trials in neuromuscular diseases (Pareyson et al., 2011a). Although the trial was negative, the natural history data from the control group can be used to assess the sensitivity to change of the outcome measures used in the study (Pareyson et al., 2011b). The primary outcome measure used as recommended by the 2005 ENMC International Workshop in CMT1A was the CMT Neuropathy Score (CMTNS) (Reilly et al., 2006). This is a validated composite scale consisting of reported symptoms (scored 0-12), physical signs (scored 0-16) and neurophysiology (scored 0-8) with a total score ranging 0-36, with <10 described as mild disease, 10-20 moderate disease and >20 severe disease (Shy et al., 2005).

Previously partly retrospective longitudinal studies have suggested progression in CMT1A patients of 0.7 points per year (Shy et al., 2008). However in the 133 patients in the placebo arm of the trial, mean change in CMTNS over two years was a 0.2 point worsening, standard deviation 2.7 points. This equates to a standardised response mean of 0.07 (0.2/2.7), indicating negligible responsiveness of this outcome measure in CMT1A patients over two years. The difference observed in per-protocol analysis of the placebo group (109 patients) was higher at 0.5 points over two years, however, the SRM is still in the minimal responsiveness range (Pareyson et al., 2011b). Interestingly, of the sub-components the symptom and signs parts of the score showed a small deterioration over two years, whereas the neurophysiology component in fact slightly improved over two years, further emphasising the difficulties in using

neurophysiology parameters as an outcome measure. Of the secondary outcome measures ankle dorsiflexion strength measured using a hand-held myometer fixed to a custom frame showed most responsiveness (SRM: 0.42 – small responsiveness), with hand grip myometry, nine-hole peg test and pain by visual analogue scale also showing SRMs in the small degree of responsiveness range.

In future trials, from Table 1-1 we can see that using outcome measures with SRM 0.2-0.4 as was seen in this study would require 100-400 patients in each arm of the study to be powered to demonstrate effectiveness of a medication which stopped progression completely. Even this in reality would signify an improvement since most of these measures would be expected to progress with normal ageing.

To increase responsiveness of an outcome measure there are two options: either increase the amount of change seen over time or reduce the standard deviation of this change by improving the reliability of the measure (Equation 1-2). This is already being applied through a new version of the CMTNS which is hoped to be both more sensitive to change and more reliable (Murphy et al., 2011a). Changes to trial design can also increase responsiveness either by increasing trial duration or selecting a homogenous group of patients likely to show the greatest change in the planned primary outcome measure. These approaches are limited by increased cost and reduced generalisability respectively.

### **1.2.5 MRI as an outcome measure**

Quantitative MRI has many characteristics that make it attractive as an outcome measure in neuromuscular diseases. MRI scanning does not involve exposure to ionising radiation and if routine precautions are followed MRI is very safe. Unlike biopsy, it is non-invasive and is able to assess any or all muscles of interest rather than a tiny sample as is the case with biopsy. Analysis is not limited to functional groups as is the case with manual strength testing or myometry – some muscles such as gracilis cannot be assessed at all by examination. Perhaps most importantly raw data is stored, which means in the context of a clinical trial that all data can be analysed by a single examiner, which eliminates inter-observer variation as a source of error. Furthermore, if new analysis methods are developed they can be applied retrospectively. None of these characteristics apply to clinical scales, patient-reported outcome measures, strength assessment or other examinations. MRI therefore is potentially powerful as an outcome measure.

MRI however is expensive and there are some technical challenges which need to be met. Although costlier than other outcome measures, in the context of the costs of a clinical trial these are manageable, and performing a clinical trial without adequate

outcome measures is an inherently futile exercise. There are many technical aspects which need to be addressed to reduce intra-scan variability such as ensuring identical patient positioning between scans, sequence parameters between scanners and analysis techniques.

### **1.3 *Principles of qualitative neuromuscular MRI***

#### **1.3.1 MRI as applied to muscle and nerve**

Magnetic resonance techniques were first applied to skeletal muscle as far back as 1965 in an examination of the effect of contraction and relaxation on transverse T2 times in the frog (Bratton et al., 1965). The use of MRI to image the central nervous system expanded exponentially in the 1980s as scanners became commercially available, with the first large series of MRI brain scans reported in 1982 (Bydder et al., 1982). Research into MRI of neuromuscular disorders resulted in many publications in the 1990s and 2000s (Koltzenburg and Yousry, 2007; Mercuri et al., 2007; Wattjes et al., 2010), however far fewer than for the central nervous system and mostly regarding qualitative assessment of pattern for diagnostic purposes. Basic MRI sequences as applied to nerve and muscle, and lower limb anatomy will now be briefly reviewed.

Imaging techniques based on x-rays such as plain film radiography and computerised tomography (CT) are reliant on a single tissue characteristic to create contrast: its propensity to absorb x-rays (radiodensity). The development of magnetic resonance imaging created contrast between tissues on the basis of entirely new tissue characteristics such as proton density, T1-recovery time, T2-relaxation time and magnetisation transfer. This allowed much greater distinction between different soft tissues and also proved more sensitive to the detection of pathology, especially in the central nervous system (McRobbie et al., 2007). A full discussion of MRI physics and sequences is beyond the scope of this thesis, but it is sufficient to understand that standard MRI sequences have been designed to create signal based on one or other of these MRI sequences: for example, a “T1-weighted” sequence is one where most of the tissue contrast is created by differences in T1-values between different tissues. Because of the way these images are generated, tissues with a short T1 value appear brighter on T1-weighted images, whereas tissues with a long T2 value appear brighter on T2-weighted images.

Translated into neuromuscular MRI this means that on T1-weighted sequences normal muscle has low signal intensity and appears dark, whereas fat, whether subcutaneous or within the muscle, has high signal intensity and appears bright. Abnormal muscle water (“muscle oedema”) cannot easily be distinguished from normal muscle on T1-weighted sequences. On standard T2-weighted sequences, normal muscle appears

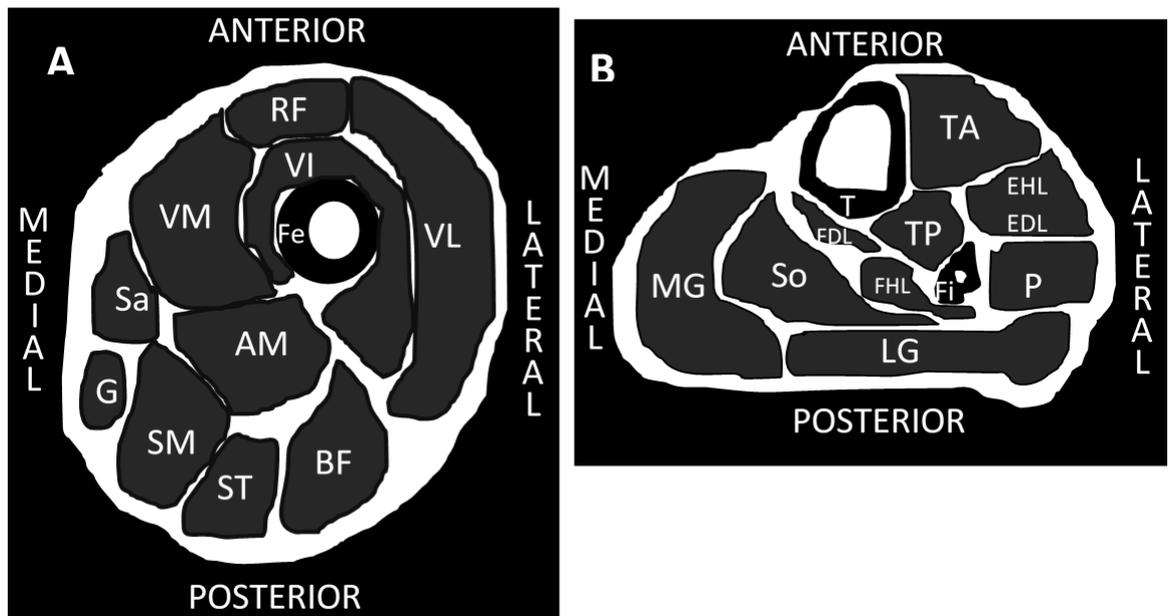
dark and fat appears bright similar to a T1 sequence, however abnormal muscle water appears bright. To better visualise abnormal muscle water on T2 sequences, an additional part is incorporated into the MRI sequence to suppress the fat signal – this is called Short-tau Inversion Recovery (STIR) – which means whilst abnormal muscle water remains brighter than normal muscle, fat is now darker than normal muscle (Koltzenburg and Yousry, 2007). This is summarised in Table 1-2. Most previously published clinical research on neuromuscular MRI has utilised T1-weighted sequences to describe the distribution and severity of fatty changes within the muscle, and STIR or other fat-suppressed T2 sequences to identify areas of muscle oedema, particularly in the inflammatory myopathies.

It should be remembered that on conventional qualitative sequences these intensities are relative, so that the onscreen appearance can be adjusted by altering brightness and contrast parameters. Tissue signal intensity can be standardised to an internal reference such as bone marrow (Ortiz-Nieto et al., 2010), or an external phantom scanned in the same field of view (Jurkat-Rott et al., 2009), but these standard clinically-used sequences are primary intended for qualitative analysis. Neuromuscular anatomy of the lower limb is outlined in Figure 1-1 for reference.

*Table 1-2: Neuromuscular MRI Tissue Characteristics*

<b>Sequence</b>	<b>Healthy Muscle</b>	<b>Muscle Oedema</b>	<b>Fat</b>	<b>Nerve</b>	<b>Injured nerve</b>
<b>T1</b>	Dark	Dark	Bright	Dark	Dark
<b>T2</b>	Dark	Bright	Bright	Dark	Bright
<b>STIR</b>	Dark	Bright	Very dark	Dark	Bright

Figure 1-1: Lower Limb Anatomy



A: Transverse section left thigh

- Bone: Fe- Femur
- Anterior compartment: RF-rectus femoris, VM-vastus medialis, VI-vastus intermedius, VL-vastus lateralis, Sa-sartorius
- Posterior compartment: G-gracilis, AM-adductor magnus, SM-semimembranosus, ST-semitendinosus, BF-biceps femoris (long head at this level)

B: Transverse section left calf

- Bones: T-tibia, F-fibula
- Anterior compartment: TA-tibialis anterior, EHL-extensor hallucis longus, EDL-extensor digitorum longus
- Lateral compartment: P-peroneal muscle group (peroneus longus at this level)
- Posterior superficial compartment: LG-lateral head of gastrocnemius, MG-medical head of gastrocnemius, So-soleus
- Posterior deep compartment: TP-tibialis posterior, FHL-flexor hallucis longus, FDL-flexor digitorum longus

### 1.3.2 MRI in inherited muscle diseases

The primary use of MRI in neuromuscular disease to date has been diagnostic, mostly for inherited muscle diseases. For this application the pattern of muscle involvement is assessed on T1-weighted lower limb sequences. Degree of fatty infiltration is usually assessed on a semi-quantitative scale, for example that described by Fischer (Fischer et al., 2008) or Mercuri (Mercuri et al., 2002a). The Mercuri scale is outlined in Table 1-3. In clinical practice, it is useful to assess which muscles are most affected or relatively spared and compare this with published patterns. This approach has proven very useful in the congenital myopathies in particular (Wattjes et al., 2010). For example in a series of MRI scans of 83 patients with one of five muscle disorders characterised, pattern analysis of the MRI identified the correct genetic diagnosis in 74

patients, with 8 scans uninformative and only one scan not in keeping with the genetic diagnosis (Mercuri et al., 2010). MRI has also been suggested as being of diagnostic utility in myofibrillar myopathies (Fischer et al., 2008) and limb-girdle muscular dystrophies (Wattjes et al., 2010). The imaging patterns have also been described in other diseases such as Duchenne muscular dystrophy or fascio-scapulothoracic muscular dystrophy (Wattjes et al., 2010), however as these diseases have a distinct clinical phenotype, MRI is probably only of additional benefit where there is diagnostic uncertainty.

*Table 1-3: Mercuri Scale*

<b>Grade</b>	<b>Description</b>
<b>0</b>	Normal appearance
<b>1</b>	Scattered small areas of increased signal
<b>2a</b>	Numerous areas of increased signal comprising less than 30%
<b>2b</b>	Numerous areas of increased signal comprising 30–60%
<b>3</b>	Washed-out appearance due to confluent areas of increased signal
<b>4</b>	End-stage appearance

*Qualitative Grading Scale of Fat Infiltration on T1-weighted MRI (Mercuri et al., 2002a)*

### **1.3.3 MRI in inflammatory muscle diseases**

MRI has been used for the inflammatory muscle diseases both diagnostically and to aid in targeting muscle biopsy to an area of inflammation (Curiel et al., 2009). For inflammatory myopathies, STIR sequences can identify areas of muscle oedema, and whole body STIR protocols can usually differentiate different types of inflammatory myopathy (Cantwell et al., 2005). Diagnosis in inflammatory myopathies is currently based on a series of clinical and pathological criteria, as unlike in inherited myopathies there can be no single gold standard test. Classification of exact diagnosis within the inflammatory myopathies is vital as whilst polymyositis and dermatomyositis respond to immunosuppressive therapy, inclusion body myositis does not, and these therapies can have severe side effects (Cox et al., 2010). A recent series of 32 patients with IBM showed distinct imaging patterns (Cox et al., 2011a) but this needs to be confirmed prospectively and any criteria validated with other conditions within the differential diagnosis for IBM.

### **1.3.4 Muscle MRI in inherited and acquired neuropathies**

There are also case reports and small series describing the pattern of muscle involvement in different subtypes of CMT (Berciano et al., 2008, 2010; Chung et al., 2008; Gallardo et al., 2006; Kim et al., 2015; Pelayo-Negro et al., 2014) but MRI for diagnostic purposes does not have an established clinical role in CMT. Secondary changes within muscles indicating both acute and chronic pathology have been

described in acquired neuropathies such as chronic inflammatory demyelinating polyneuropathy (Sinclair et al., 2012a).

The changes seen within muscles of patients with neuropathic conditions are in general similar to those seen in diseases where the primary pathology is myopathic. That is, chronic neurogenic fatty atrophy is seen as a hyperintense signal on T1w images with reduced muscle size, whilst acute neurogenic changes are seen as hyperintensity within muscles on STIR sequences. Whilst this lack of specificity is a limitation in terms of differentiating neuropathic from myopathic weakness for diagnostic purposes, it is potentially a great strength for quantitative assessment of disease severity and progression as the same methods can be utilised regardless of disease pathogenesis.

#### **1.4 *Quantitative neuromuscular MRI***

The studies described above have utilised standard MRI sequences, which are considered qualitative and are well suited to diagnosis. Semi-quantitative scales, such as that of Mercuri for fatty infiltration described in table 1.3, may be utilised for an outcome measure but are not ideal for this purpose as there are only six discrete points on this scale, so they lack sensitivity to measure change over time, especially in a gradually progressive disease. Furthermore, their assessment is subjective which is likely to increase both inter-observer and intra-observer variation in scoring. Some researchers have quantified a T1 image by calculating the area of muscle with hyperintensity and the area of muscle with normal intensity, as applied to patients with spinal muscular atrophy (Sproule et al., 2011a). Whilst this gives a result which is a continuous variable rather than an ordinal scale, so would be potentially sensitive to change over time, it is still utilising an inherently non-quantitative sequence with all the sources of variation noted above. Despite excellent test-retest reliability reported (Sproule et al., 2011a) they were unable to show significant change in thigh muscle volume over six months using this technique (Sproule et al., 2011b), although other researchers have shown significant muscle volume loss in patients with diabetic neuropathy with this method (Andreassen et al., 2009).

##### **1.4.1 T1 and T2 mapping**

The alternative is to attempt to measure the specific tissue characteristics responsible for the signal in conventional MRI. The simplest approach is to utilise MRI sequences which estimate either the T1-recovery times or the T2-relaxation times of tissue. In healthy volunteers T1-recovery time is about 1300ms for muscle and 350ms for fat, whilst T2 relaxation times are 40ms and 180ms respectively (see section 3.2.1). An image can then be constructed where the brightness of each pixel relates directly to

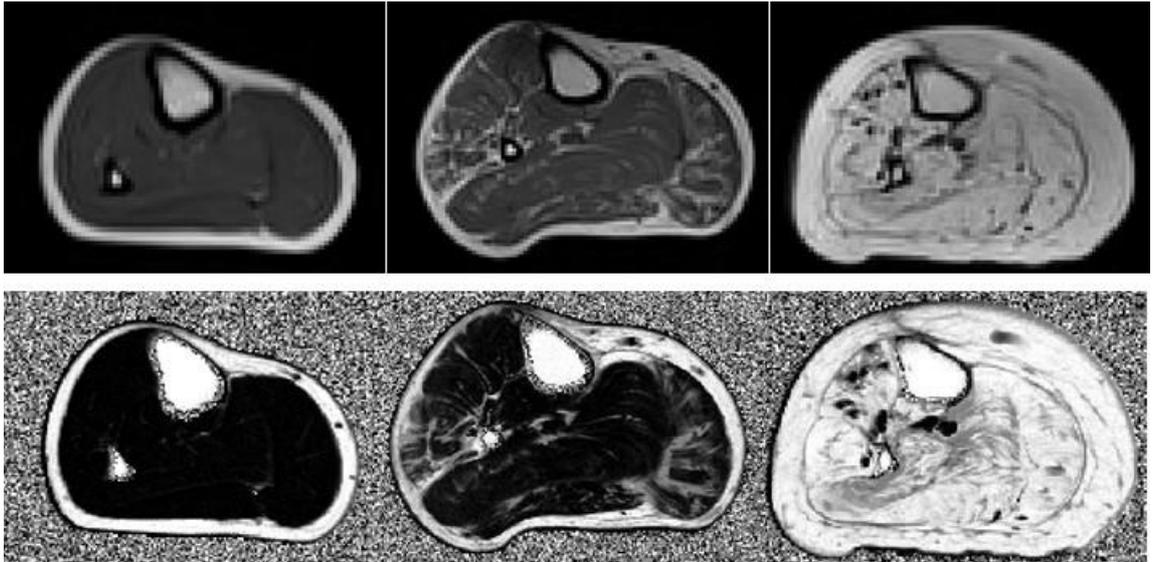
the estimated T1-relaxation time: a “T1-map”. There are many different methods to estimate T1-relaxation time, but the general principle is to measure signal intensity from a volume of tissue repeatedly whilst varying a sequence parameter (for example flip angle), then calculating the T1 value which fits the observations. Similarly, T2 time is estimated by fitting a model based on signal intensity measured whilst varying echo times.

These maps can then be analysed in a number of ways to create potential outcome measures. Mean values can be calculated of a region of interest (ROI) drawn within a muscle, comprising a whole muscle on a single slice, comprising a whole muscle across multiple slices or averaged against multiple muscles. Alternatively, histogram analysis will produce the mode value, which can be applied to a whole slice using automated analysis which doesn't require manual ROIs to be drawn. These techniques have been applied to measure T2-relaxation times in muscle of patients with Duchenne muscular dystrophy. These techniques have been applied to measure T2-relaxation times in muscle of patients with Duchenne muscular dystrophy (Huang et al., 1994; Kim et al., 2010a), juvenile dermatomyositis (Maillard et al., 2004a) and amyotrophic lateral sclerosis (Bryan et al., 1998) and T1-recovery times in DMD (Huang et al., 1994). These methods can be further developed by the addition of fat-water separation or fat suppression methods, for example the IDEAL-CPMG MRI pulse sequence (Janiczek et al., 2011). This combines fat-water separation and T2 relaxation time quantification in a single acquisition, allowing the T2 of the water component to be distinguished from that of fat, something that has not been possible up to this point.

#### **1.4.2 Fat quantification**

Two additional quantitative MRI tissue parameters have promising data in respect to muscle pathology. The three-point Dixon fat/water separation method involves taking three acquisitions which can be later processed to create a fat-only image and a water-only image (Glover and Schneider, 1991). These may then be used to generate a fat-fraction map where the relative proportion of fat to water within each voxel is calculated as a percentage (see Figure 1-2). This can be used to create a visual image or analysed using ROI or automated techniques in a similar fashion to T1 or T2 quantitative techniques described above. This technique has been used extensively in the assessment of non-alcoholic steatohepatitis (fatty liver) and within the neuromuscular field in DMD (Wren et al., 2008), where it has been proposed as a better outcome measure than manual muscle testing or dynamometry.

Figure 1-2: Three-point Dixon



Fat maps below corresponding T1-weighted conventional transverse MRI images through the mid-calf of a healthy volunteer, moderate and severe CMT1A patients (left to right). Black represents voxels with 0% fat, white 100% fat.

### 1.4.3 Magnetisation transfer imaging

Whilst fat deposition is considered the end-stage of muscle pathology whether due to a myopathic or neurogenic process, earlier changes may be observed through analysing the distribution of water within the muscle tissues. The traditional approach to this in qualitative neuromuscular MRI is to utilise a STIR sequence which is T2-weighted with fat suppression such that “muscle oedema” appears bright where normal muscle tissue and fat appear dark. Some researchers have made a semi-quantitative analysis of this by measuring signal intensity with reference to a saline phantom (Jurkat-Rott et al., 2009), but it isn’t possible to perform true quantitative analysis on the STIR sequence.

Another approach is to utilise the principle of magnetisation transfer to analyse the muscle water distribution. This technique is based on the presence of two pools of water within any tissue – a pool which is bound to macromolecules known as the bound pool and water which moves freely known as the free pool. The source of the signal in conventional MRI is the hydrogen atoms within the water molecules in the free pool, whereas the hydrogen atoms in the water molecules of the bound pool are effectively invisible to conventional MRI due to a wide band of resonance frequencies. However, this pool can be indirectly assessed by applying a magnetic pulse to saturate this pool prior to obtaining an MRI acquisition. This results in lower signal intensity than one acquired without the saturation pulse and this reduction may be expressed as the “magnetisation transfer ratio”, the magnitude of which is determined by the proportion of water molecules in the bound pool. The MTR of healthy muscle is around 40 percentage units (p.u.) and it has been shown to be reduced in skeletal muscle in a

variety of conditions where there is an alteration to normal muscle structure, for example in limb girdle muscular dystrophy (McDaniel et al., 1999) and in extra-ocular muscles in thyroid eye disease (Ulmer et al., 1998). This technique may show abnormalities earlier than those due to fatty deposition, and these may be potentially reversible with appropriate therapy.

## **1.5 *Aims of thesis***

In summary, neuromuscular diseases are in desperate need of an outcome measure which is reproducible and sensitive to change, and quantitative MRI has the characteristics necessary to fill this role. The overall aim was to assess the utility of MRI as an outcome measure in neuromuscular diseases. This was undertaken in three phases. First, the fundamental properties of lower limb muscles quantified with MRI were determined in healthy volunteers, including inter-muscle variation, inter-subject variation due to demographic variation, test-retest reliability, inter-observer reliability, relationships between quantitative MRI parameters and between qualitative and quantitative assessment. These are presented and discussed in detail in chapter 3. These methods were then applied together with clinical assessment and myometry in patients with IBM and CMT, primarily to demonstrate the validity of quantitative MRI in patients with chronic neuromuscular diseases. This is presented in chapter 4. Finally, longitudinal assessments were performed to assess the responsiveness of quantitative MRI measurements in neuromuscular diseases. These are presented in chapter 5. Undertaking these three steps has identified valid, reliable responsive outcome measures, ready for use in clinical trials of neuromuscular diseases.

## **2 Methods**

### **2.1 Background**

#### **2.1.1 Development of novel outcome measures**

The establishment of a novel outcome measure may be considered as a five step process: 1) basic scientific research to develop a new method of measurement; 2) assessment of the outcome measure in healthy volunteers to assess reliability; 3) cross-sectional assessment of the outcome measure in patients to assess validity; 4) longitudinal assessment of the outcome measure in patients to assess responsiveness; 5) application of the outcome measure in an interventional clinical study. These same steps can be applied to any outcome measure: from a simple quality of life questionnaire to the most complex MRI measurement technique.

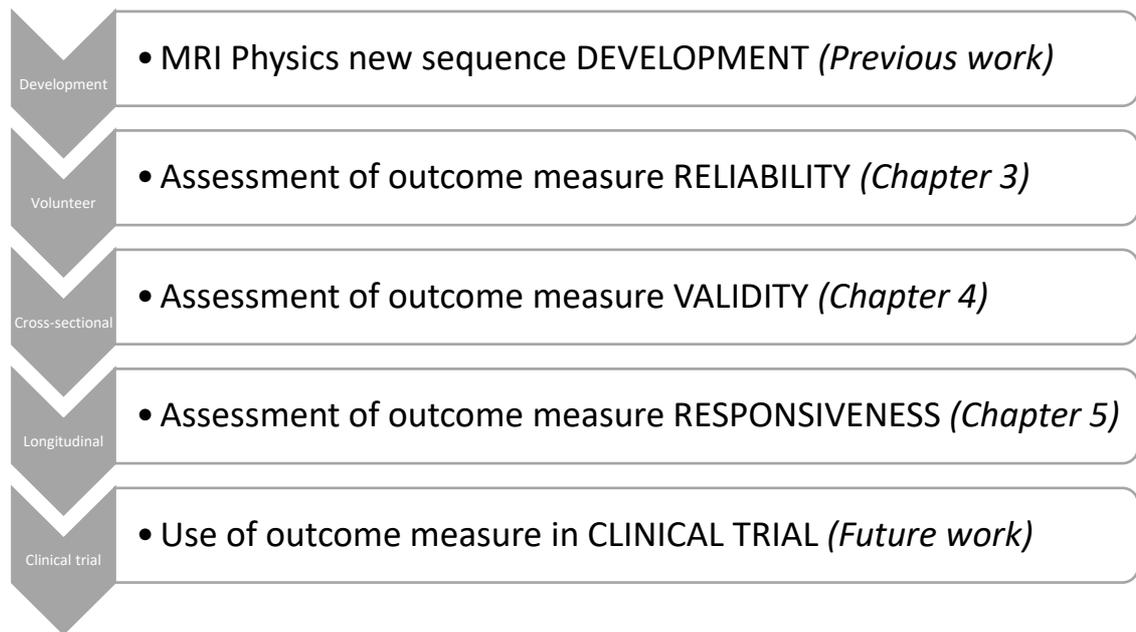
In practice the process may not be as linear as suggested in the schematic for MRI outcome measure development (Figure 2-1). For example, results obtained during assessment in healthy volunteers may lead to further refinement of the MRI methods. Reliability should be assessed using patient data as well as healthy control data. Validity should also be demonstrated through correlation of clinical and MRI data longitudinally, and ultimately responsiveness is best demonstrated in an interventional study. However, failure to adequately complete one of the first four steps before application in a clinical trial can be disastrous.

Quantitative MRI sequences utilised in skeletal muscle imaging of neuromuscular diseases were initially developed for use in other tissues such as T2 mapping of the brain and Dixon fat-water separation method for the liver. These have been applied to skeletal muscle previously by many centres including the MRC Centre for Neuromuscular Diseases, in a pilot study prior to the commencement of this thesis (Sinclair et al., 2012a). The purpose of this thesis is to systematically move through the second, third and fourth steps, so that quantitative MRI might be considered ready for application in a clinical trial.

The study was thus designed in three phases:

1. Reliability study in healthy volunteers
2. Cross-sectional study in patients and matched healthy controls
3. Longitudinal study in patients and matched healthy controls

Figure 2-1: Outcome Measure Development Pipeline



The major purpose of the first phase was to assess reliability of quantitative MRI measurements with assessment of both test-retest and inter-rater reliability. Together with cross-sectional data from the healthy volunteers, phase 1 data also determined the normal variation of quantitative MRI measurements in healthy volunteers, both between muscles and between subjects. The major purpose of phase 2 was to demonstrate outcome measure validity by correlating MRI and clinical measures, whilst also defining the range and patterns of abnormalities seen in patients. Finally, the major aim of phase 3 was to assess outcome measure responsiveness, through longitudinal assessment of different MRI biomarkers.

### 2.1.2 Disease group selection

Previous studies of quantitative MRI in neuromuscular diseases have mostly focussed on a single disease group and only occasionally a mixed disease group (see 4.2.2). Most have focussed on myopathic conditions and the only previous paper of quantitative muscle MRI in Charcot-Marie-Tooth disease (CMT) was the pilot study using MTR from our group (Sinclair et al., 2012a). In the present study, we wished to demonstrate the wide applicability of quantitative MRI and hence chose two very different conditions: CMT and inclusion body myositis. They both are characterised by lower limb weakness, but are otherwise very different, respectively: inherited/acquired, neurogenic/myopathic, very slowly progressive/moderately progressive, symmetrical/asymmetrical, with predominant calf/thigh weakness. Thus whilst maintaining homogeneity by including only two specific neuromuscular disease groups; between them they characterised a broad range of neuromuscular diseases.

### **2.1.2.1 Inclusion body myositis**

IBM is an acquired inflammatory myopathy with onset usually over 50 years of age with a male predominance and the quadriceps being the predominant muscle affected in the lower limbs. There are currently no treatments with proven efficacy despite numerous trials (Hilton-Jones et al., 2010). Disease progression is moderate with a median duration from symptom onset to wheelchair use of 15 years (Benveniste et al., 2011; Cortese et al., 2013). These characteristics would be predictive of greater evidence of active disease on MRI (T2/STIR/MTR), with evidence of disease progression likely over 12 months with a moderately responsive outcome measure. There are a number of proposed clinical trials of potential treatments in IBM, such as arimoclomol (Dimachkie et al., 2014), so a sensitive outcome measure is needed.

### **2.1.2.2 Charcot-Marie-Tooth disease 1A**

Charcot-Marie-Tooth disease (CMT) is an umbrella term for the hereditary motor and sensory neuropathies with a combined prevalence of around 1:2500 and over 70 causative genes identified. The most common type is CMT1A representing 50% of all CMT with progressive distal weakness from childhood. Progression is extremely slow with the vast majority of patients never losing ambulation (Reilly et al., 2011). Recent trials have been hampered by insensitive outcome measures (see 1.2.4) (Pareyson et al., 2011a). These characteristics would be predictive of a lesser degree of active pathology (denervation) on MRI with only a highly responsive outcome measure likely to show significant change over 12 months. Despite the unsuccessful vitamin C trials, there are a number of other potential therapies (Patzkó and Shy, 2010) which will require rigorous assessment in clinical trials. Highly responsive outcome measures would be needed to allow adequately powered studies with a feasible number of participants.

### **2.1.3 Use of healthy controls**

Some quantitative MRI studies in neuromuscular diseases included a healthy control group (Gloor et al., 2011; Hiba et al., 2012; Huang et al., 1994; Kan et al., 2009; Maillard et al., 2004b; McDaniel et al., 1999; Sinclair et al., 2012a), whilst many have not (Gaeta et al., 2011; Kim et al., 2010a; Willis et al., 2013; Wren et al., 2008). As in the first phase of this study, healthy controls may be used to assess inter-scan reliability, but healthy controls serve additional vital functions in both the cross-sectional and longitudinal patient study phases.

In a cross-sectional study, the inclusion of healthy controls allows direct comparison of patient data with values in healthy subjects. As can be seen from the literature review in 3.2.1, reported values of muscle MRI measurements such as T1, T2, fat fraction and

MTR vary significantly between studies, due to differences in field strength, scanner manufacturer and model, sequence parameters, and analysis methods, so it is not possible to compare results in patients to those of healthy controls in the published literature. Even to compare results from an earlier volunteer study at the same site may be problematic. There may be changes in scanner hardware or software outside of the control of the investigators, such as that caused by the scanner software upgrade in this study (see 2.4.4.3.1). This may introduce a systematic bias between the healthy controls scanned at an earlier time point and patients in a new study, or in a longitudinal study between baseline and follow-up scans. Thus for quality control reasons, healthy controls are essential to any quantitative MRI study.

There is an additional crucial consideration in longitudinal assessment of outcome measures in neuromuscular diseases: the confounding effect due to healthy ageing. The 5 year natural history study by Verhamme and colleagues (Verhamme et al., 2009) exemplifies this point. This study looked at a range of clinical, myometric and neurophysiological measurements in both patients with CMT1A and healthy controls. Just considering the patient data, they identified two measurements with moderate responsiveness (SRM = -0.60) over 5 years: three-point grip strength and ulnar motor amplitude on neurophysiology. However, their healthy control group also deteriorated on both these measures, and the change in patients was not significantly greater than in healthy volunteers, leading the authors to infer that these changes were due to the effects of ageing rather than disease progression. Whilst the authors' conclusion based on these results that there is no progression in CMT1A in adulthood beyond that resulting from normal ageing is at odds with clinical experience, this study demonstrates the crucial role of healthy controls in CMT1A natural history studies. Whilst in IBM the rate of progression is greater, the patient group is also significantly older so the confounding effects of ageing will likely also be greater.

#### **2.1.4 Sample size**

For the healthy volunteer study,  $n = 15$  was chosen for test-retest assessment. This is an adequate number to provide estimates of outcome measure reliability; previous studies have utilised 10 healthy controls for this purpose (Pfirrmann et al., 2004; Smeulders et al., 2010). For the longitudinal study, 20 patients provide 80% power to detect significance of a mean change of 0.66 SD units, assuming 10% dropout using a paired t-test and a 5% significance level. The patient data reported here is the start of a longer study with five years' follow-up, with correspondingly greater power. With extremely limited longitudinal quantitative muscle MRI data available in healthy subjects, we chose to include a healthy control group of equal size for both patient groups.

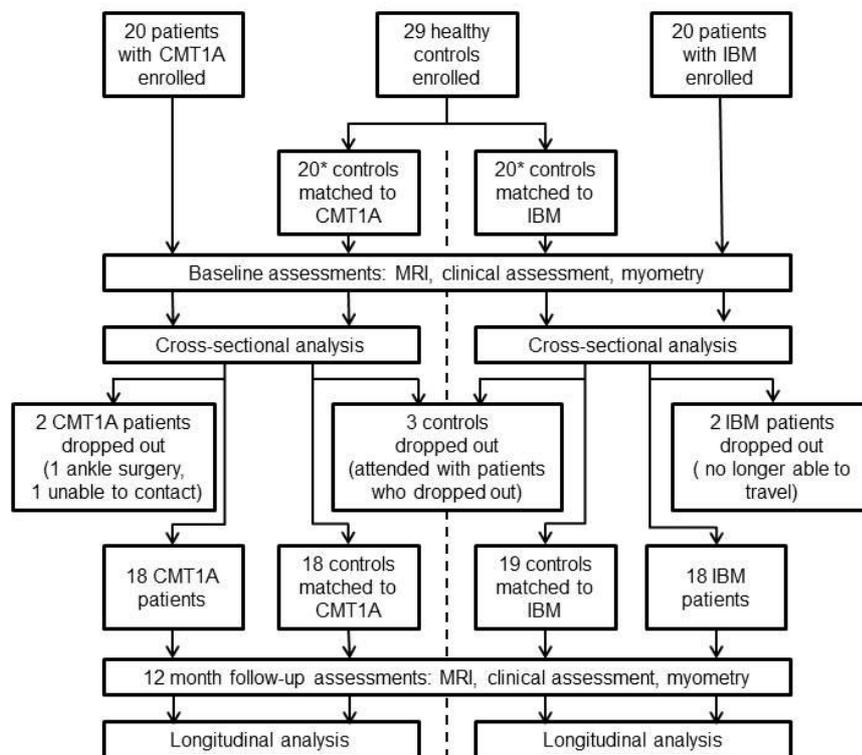
## 2.1.5 Ethical framework

The study was approved by the local research ethics committee (see acknowledgements to Dr Sinclair, Dr Machado and Dr Miller). All participants were provided with a written information sheet with regard to the study and provided signed informed consent after at least 24 hours consideration.

## 2.2 Study design

Phase 1 was a prospective cross-sectional observational study in 18 healthy volunteers, with 15 of these able to undertake repeat imaging using an identical protocol at a two week interval. Healthy volunteers enrolled in this phase undertook a screening history and examination to rule out neuromuscular disease, and completed standard MRI safety screening but did not undertake any other clinical assessments such as myometry. Healthy volunteers in phase 1 were not enrolled as controls for subsequent study phases. The patient study (phases 2 and 3) utilised a prospective longitudinal observational study design with two sets of assessments at a 12 month interval.

Figure 2-2: Flow chart of subject assessments and drop out for phases 2-3



\* 11 controls were common to both disease control subgroups

### 2.2.1 Subject recruitment

Patients meeting the inclusion criteria attending the inherited neuropathy or muscle clinic at the National Hospital for Neurology and Neurosurgery or the IBM research clinic at the MRC Centre for Neuromuscular Diseases were invited to participate. Healthy controls were recruited from hospital and research staff, and from the spouses, friends and relatives of the patients (blood relatives of CMT1A patients only included if genetic testing negative). Subjects underwent initial assessments between January 2010 and July 2011. The inclusion criterion for CMT1A patients was genetic confirmation of chromosome 17p11.2 duplication, whilst for IBM patients the inclusion criterion was meeting pathologically or clinically definite by MRC IBM diagnostic criteria (Hilton-Jones et al., 2010). Exclusion criteria related to concomitant diseases and MRI safety, listed in Figure 2-3.

Sufficient controls were required to allow a subset of 20 to be age and gender matched to both the CMT1A patient group and the IBM patient group. As the demographics of the two patient differed (IBM being a disease predominantly affecting men over 50), recruitment of a total of 29 control subjects was required to allow arrangement into two groups of 20, designated control (IBM) and control (CMT1A), 11 of which were controls to both groups. Age, gender, height and weight showed no significant differences between patients and their matched control subsets.

*Figure 2-3: Exclusion criteria for all subjects*

- Concomitant neuropathy/myopathy
- Very advanced disease state that precludes travelling
- Severe cardiovascular, renal, or other end-stage-organ-disease states or any other major comorbidities (e.g. any active malignancy, definite cognitive impairment, psychiatric disease, heart or lung failure, orthopaedic or rheumatologic disorders)
- Pregnancy and active nursing (breast feeding)
- Inadequate contraception in women of child bearing age.
- Contra-indication to MRI (metallic fragments, clips or devices in the brain, eye, spinal canal, etc; magnetically activated implanted devices, such as cardiac pacemakers, insulin pumps, neurostimulators and cochlear implants; claustrophobia)
- Existing radiculopathy or lower back pain
- Surgery to feet within 12 months of beginning of study or during the study

### 2.3 Clinical assessments

Patients and controls underwent clinical assessment, bedside strength testing and myometry at each baseline and follow-up visit. Patients also performed a quality of life questionnaire and were scored on disease specific scales.

### **2.3.1 Disease specific scales**

A number of outcome measures have been utilised in natural history studies and clinical trials in CMT1A and IBM. The CMT neuropathy score (CMTNS) is a 36 point scale, scored 0-4 on 9 items, with 3 history items, 4 examination items and 2 neurophysiology items and has been shown to be a reliable and valid marker of disease severity (Shy et al., 2005). In CMT1A the CMTNS was used as the primary outcome measure in the completed adult vitamin C trials (Lewis et al., 2013; Pareyson et al., 2011a) based on the recommendation of the 2005 ENMC International Workshop (Reilly et al., 2006). The CMTNS performed poorly in both these studies, with the placebo group showing minimal change in one study (Pareyson et al., 2011a), and paradoxical improvement in the other (Lewis et al., 2013). Neurophysiology measurements are a component of this score, and were used as the primary outcome measure in the paediatric study in vitamin C (Burns et al., 2009) but showed minimal responsiveness in the placebo arms of all three studies, and showed no significant deterioration compared with healthy controls in a large 5 year natural history study (Verhamme et al., 2009). Standard neurophysiology techniques have not been proposed as an outcome measure in inclusion body myositis. For these reasons, plus the added burden on participants, additional neurophysiology wasn't performed as part of assessments in this current study. The CMTNS without neurophysiology: the CMT Examination Score (CMTES) was therefore used as the main overall clinical severity measurement in CMT1A patients, using a version modified to improve sensitivity just prior to the commencement of the study (Murphy et al., 2011b).

The IBM functional rating score is an IBM specific scale adapted from the ALS FRS (Jackson et al., 2008). It is reliable, valid, and was the most sensitive measure of change in a study of high dose interferon-1a in IBM (Jackson et al., 2008). Although physician-administered, it could easily be adapted to be a patient reported outcome measure. For consistency, it was administered during the study by the thesis author in all cases.

### **2.3.2 Myometry**

Patients and controls underwent detailed lower limb myometry on a HUMAC NORM dynamometer (CSMi, Massachusetts, USA). The myometry assessment either occurred following the MRI or with at least a 60-minute gap if preceding to minimise any potential MRI effects secondary to physical activity.

All but two initial assessments and all follow up assessments were performed by the thesis author. Myometry was performed according to Table 2-1. Knee extension, knee flexion, ankle dorsiflexion, ankle plantarflexion, ankle inversion and ankle eversion

were assessed bilaterally using both isometric and isokinetic protocols and the maximum torque in Nm recorded for analysis. Isometric assessments consisted of four attempts of 3 seconds duration with 10s interval of which the best attempt was selected. For the isokinetic assessments, following a practice run and 10s interval, three successive movements through full range were performed and the highest value obtained selected. The machine setup was recording at first visit using the included software, which was then retrieved to allow identical set-up on repeat testing.

*Table 2-1: Myometry protocol*

Joint	Type	Movement	Angle
Knee (right then left)	Isometric	Extension°	45°
		Extension	90°
		Flexion	45°
		Flexion	90°
	Isokinetic	Extension/Flexion	60°/s
		Extension/Flexion	120°/s
Ankle (left then right)	Isometric	Plantarflexion	10°
		Dorsiflexion	10°
	Isokinetic	Plantarflexion/Dorsiflexion	60°/s
Ankle (right then left)	Isometric	Eversion	0°
		Inversion	0°
	Isokinetic	Inversion/Eversion	60°/s

### **2.3.2.1 Myometry analysis**

All measurements are stored within the HUMAC NORM system, and were checked for artefactual values before being exported into IBM-SPSS for analysis. An additional overall strength for each movement was calculated by taking the mean of all assessment methods listed in Table 2-1.

### **2.3.3 MRC strength grading**

The following movements were assessed:

- Neck: flexion and extension
- Upper limbs: shoulder abduction, elbow extension, elbow flexion, wrist extension, wrist flexion, finger extension, forefinger abduction, little finger abduction, thumb abduction, long finger flexors, short finger flexors
- Lower limbs: hip flexion, hip extension, hip abduction, hip adduction, knee flexion, knee extension, ankle dorsiflexion, ankle plantarflexion, ankle eversion, ankle inversion, big toe extension

Muscle strength was assessed using a modified MRC scale (O'Brien, 2000) with 5: Normal strength, 5-: Barely detectable weakness, 4+: Gravity and moderate to maximal resistance, 4: Gravity and moderate resistance, 4-: Gravity and minimal resistance, 3: Full range of motion against gravity only, 2: Movement when gravity is eliminated, 1: Flicker of movement seen or felt, 0: No movement. Upper limb/neck and lower muscle scores were summed to obtain a total upper limb/neck and total lower limb score for each subject. For this purpose, 5- was scored as 4.75, 4+ as 4.25 and 4- as 3.75. The maximum score obtainable was 120 for upper limb/neck and 110 for lower limb.

#### **2.3.4 Other assessments**

Additional clinical assessments performed in both patient groups were: medical history, neurological examination, and SF-36 quality of life questionnaire (Ware and Sherbourne, 1992). Upper limb strength tests and upper-limb functional tests were not performed, due to the lower limb focus of the study. The 6 minute-walk test has subsequently become a commonly used measure in clinical trials of muscle diseases, being used to the primary outcome measure in a licensing trial of enzyme replacement therapy in Pompe disease (van der Ploeg et al., 2010). It has not been validated in CMT1A, and is highly unlikely to be responsive as CMT1A patients maintain ambulation long term. It has not been validated in IBM, and although may prove useful, is likely to be prone to the same non-linear relationship with disease progression due to a sudden drop off when patients become non-ambulant (McDonald et al., 2010).

### **2.4 MRI protocol**

The protocol was primarily developed by the MR Physics team involved in the project: Dr Sinclair and Dr Thornton (see acknowledgements).

#### **2.4.1 Hardware**

Subjects were examined lying feet-first and supine at 3T (TIM Trio, Siemens, Erlangen, Germany) using a multi-channel peripheral angiography coil (Siemens 'PA Matrix') and 'spine matrix' coil elements.

#### **2.4.2 Subject positioning**

Before scanning, the distance between the anterior superior iliac spine and the superior border of the patella was measured and thigh-level imaging volumes were centred one third of this distance above the patella superior border. For phase 1, calf-level imaging volumes were centred on the point of widest lower leg circumference. For phases 2 and 3 calf-level imaging volumes were centred below the tibial tuberosity

by one quarter of the total distance from the tibial tuberosity to the lateral malleolus. The derived distances at baseline were recorded and used for block positioning on follow up imaging.

### **2.4.3 Anatomical coverage**

Lower limb muscles were chosen as the region for study as they are a key site of pathology in both CMT1A (lower calf) (Reilly et al., 2011) and IBM (quadriceps) (Hilton-Jones et al., 2010) and weakness in these areas is a key cause of disability in these patient groups. Furthermore, lower limb imaging has practical advantages over dedicated upper limb imaging: both limbs may be imaged simultaneously, lowering scanning times, and lower limb imaging is in our experience more comfortable for participants.

Both limbs were scanned within the FOV. Axial-slice matrices and fields of view (FOV) were 256x128 and 400x200mm (410x205mm in some phase 1 subjects) for thigh- and 256x120 and 400x188mm for calf-level images, except for FF acquisitions where matrices were 512x256 and 512x240 pixels respectively.

### **2.4.4 MRI sequences**

The total acquisition time was less than 60 minutes. The following sequences were performed: T1-weighted and short-tau-inversion recovery (STIR) qualitative imaging, fat fraction measurement using three-point-Dixon technique (Glover and Schneider, 1991), T1 relaxometry by DESPOT method with B1 correction (Deoni et al., 2003), pseudo-T2 relaxometry from dual-contrast turbo-spin-echo images, magnetisation transfer ratio derived from two 3D-FLASH images with and without an MT pre-pulse with B1 correction (Sinclair et al., 2012b).

#### **2.4.4.1 Fat fraction measurement**

For Dixon FF measurements (Glover and Schneider, 1991), three 2D gradient-echo acquisitions were performed with parameters (TE1/TE2/TE3=3.45/4.60/5.75ms, TR=100ms, flip angle=10°, bandwidth 420Hz/pixel, NEX=4, 10 x 10mm slices with 10mm gap, 512x256 matrix (thigh), 512x240 matrix calf, iPat=2). Phase unwrapping was performed using PRELUDE (FSL, FMRIB, Oxford) (Smith et al., 2004) and after fat (F) and water (W) image decomposition, FF calculated as  $FF = 100\% \times F/(F+W)$ .

The TE=3.45ms image was used for the region of interest (ROI) placement and as a reference for inter-method image interpolation and registration (calf level only) using FLIRT (FSL, FMRIB, Oxford), such that that the same ROIs could be applied to extract data from all maps.

#### 2.4.4.2 T<sub>1</sub>-Relaxometry

DESPOT-1 (Deoni et al., 2003) T<sub>1</sub>-mapping used three 3D fast low-angle shot (3D-FLASH) images S<sub>1,2,3</sub> with nominal α<sub>1,2,3</sub> of 5, 15 and 25°, TR/TE=23/3ms, and BW=440Hz/pixel acquired in a single, non-selective slab with 80 x 5mm longitudinal phase-encoded partitions. Flip-angles were corrected using B<sub>1</sub> maps obtained as below and T<sub>1</sub> calculated according to Deoni (Deoni et al., 2003).

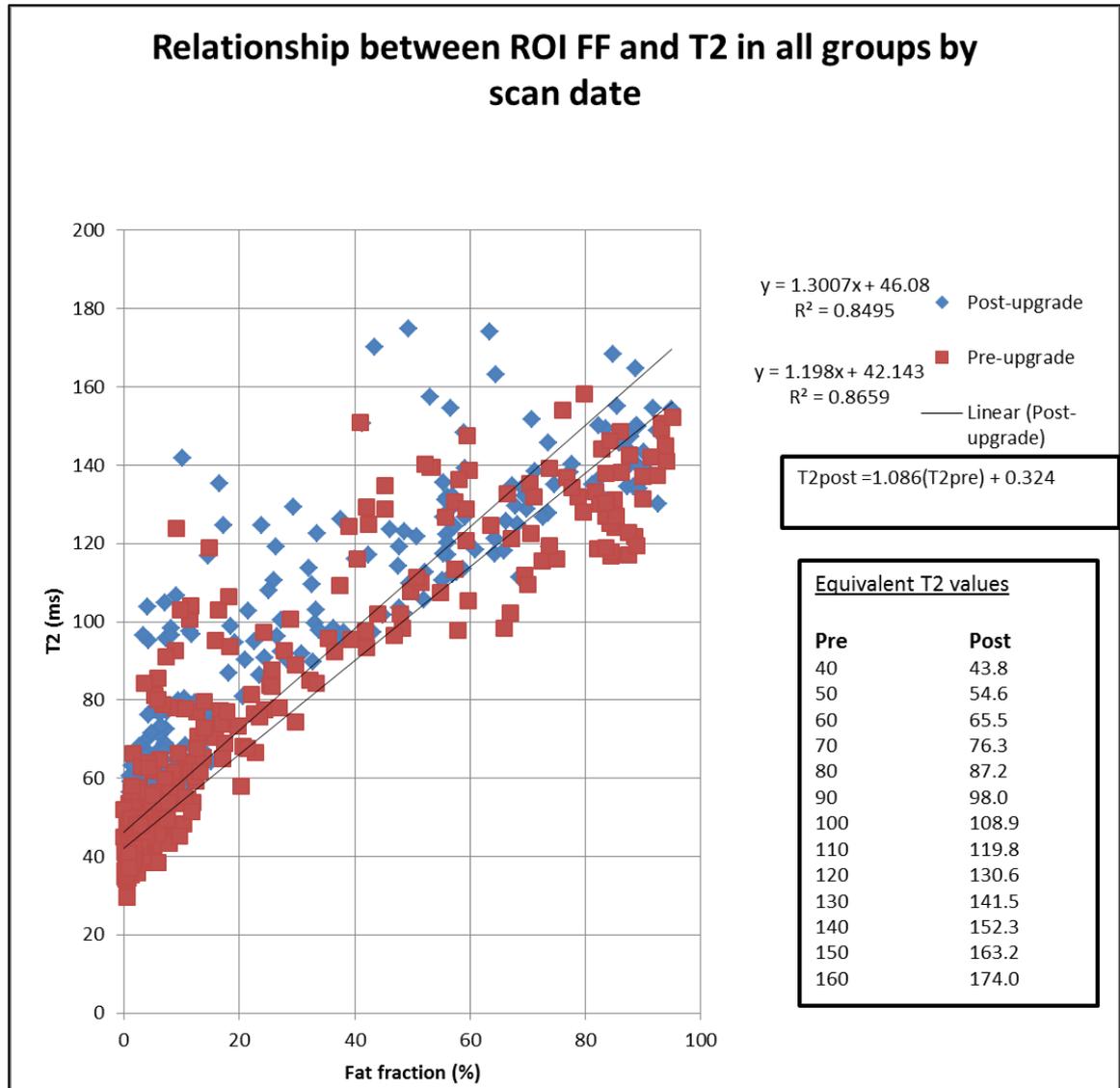
#### 2.4.4.3 T<sub>2</sub>-Relaxometry

Dual-contrast turbo-spin-echo (TSE) images (6500/13/52ms or 6500/16/56ms; 10x10mm slices with 10mm gap, iPat=2, BW=444Hz/pixel, refocusing flip angle 180°, NEX=1, 6/8 *k*-space sampling, 256x128 matrix (thigh), 256x120 matrix (calf)) were acquired. Pseudo-T<sub>2</sub> was calculated from the respective pixel intensities I<sub>TE1</sub> and I<sub>TE2</sub> from the TE<sub>1</sub> and TE<sub>2</sub> images as  $T_2 = \frac{TE_2 - TE_1}{\ln(I_{TE1} / I_{TE2})}$ .

##### 2.4.4.3.1 Software upgrade effects on T<sub>2</sub>

The different echo times were the result of altered sequence timing constraints following a routine scanner software upgrade which occurred after 54 baseline and 6 follow-up scans had been completed. Analysing control values pre- and post-upgrade suggested a systematic bias between pre- and post-upgrade T<sub>2</sub> values (see **Error! Not a valid bookmark self-reference.**). By comparing the observed relationship between FF and T<sub>2</sub> measurements pre- and post-software upgrade, a correction equation was determined separately for thigh muscles (corrected T<sub>2</sub> = 1.0606 x post-upgrade T<sub>2</sub> + 1.1522) and calf muscles (corrected T<sub>2</sub> = 1.0933 x post-upgrade T<sub>2</sub> – 0.0245). These corrections were applied to all post-upgrade T<sub>2</sub> measurements to ensure pre- and post-upgrade T<sub>2</sub> measurements were comparable. Parameters for the other quantitative sequences were not affected by this upgrade, and analysis of the control values pre- and post-upgrade indicated no systematic bias in these values was introduced by the software upgrade.

Figure 2-4: Analysis of Upgrade effects on T2



#### 2.4.4.4 B<sub>1</sub> Mapping

Separate TSE images (TR/TE=7000/11ms, 128x64 matrix, 40 contiguous 10mm slices, BW=429Hz/pixel, 1/2 *k*-space sampling) yielded image intensities  $V_1$  and  $V_2$  acquired with nominal excitation  $\alpha_1$  and  $\alpha_2$  of 60° and 120°. B<sub>1</sub> deviation was mapped according to  $B_{1Dev} = \arccos(V_2 / 2V_1) / \alpha_1$  (Stollberger and Wach, 1996).

#### 2.4.4.5 Magnetization transfer ratio

MTRs were calculated from two 3D-FLASH images with ( $M_1$ ) and without ( $M_0$ ), an MT pre-pulse (500° amplitude, 1200Hz offset, 10ms duration) (TR/TE=65/3ms or 68/3ms,  $\alpha=10^\circ$ , BW=440Hz/pixel, NEX=1, 6/8 *k*-space sampling, iPat=2, 40x5mm longitudinal phase encoding partitions, 256x128 matrix (thigh), 256x120 matrix (calf)) according to  $MTR = (M_0 - M_1) / M_0 \times 100$  percentage units (p.u.). Percentage units were used by convention to avoid ambiguity with fractional change expressed as a percentage

(Barker et al., 1996). MTR maps were RF-inhomogeneity corrected using the  $B_1$  maps according using a mean-over-all-subjects  $B_1$  inhomogeneity correction factor of  $k = 0.0085$  (Sinclair et al., 2012b).

#### **2.4.4.6 T1 weighted imaging**

Standard T1-weighted images were acquired with a turbo-spin-echo readout prior to commencing the quantitative protocol (TR/TE=671/16ms, 10 slices, 10mm thickness, 10mm slice gap, 256x192 matrix, iPat acceleration of 2, 444 Hz/pixel bandwidth (BW), TSE factor=3, refocusing flip angle (fa) 130°, NEX=2, acquisition time (TA) = 45s).

#### **2.4.4.7 Short-tau-inversion-recovery imaging**

STIR (TR/TE/Inversion Time = 5500/56/220ms, NEX=1, flip angle 180°, parallel imaging factor (iPat)=2) imaging was performed (10x10mm slices, 10mm gap, 256x128 matrix (thigh), 256x128 matrix (calf)).

### **2.5 Qualitative image analysis**

Analysis of apparent fatty infiltration on the T1 weighted images was performed by grading using the 6-point Mercuri scale (Mercuri et al., 2002b) (0=normal, 1=mild fatty streaks, 2a=early confluence, 2b=fatty infiltration 30-60%, 3=fatty infiltration >60%, 4=complete fat replacement) and for STIR hyperintensity using a three-point scale (0=none, 1=mild, 2=marked) by a radiologist with 4 years post-speciality experience (Dr Fischmann, acknowledgements). The observer was blinded to subject group (IBM/CMT1A/control). Repeat scans were graded at a later time-point without reference to baseline scans. The same muscles were assessed as in the quantitative analysis below, but the entire muscle volume imaged was assessed, rather than just a single slice.

### **2.6 Quantitative analysis**

#### **2.6.1 Slice selection**

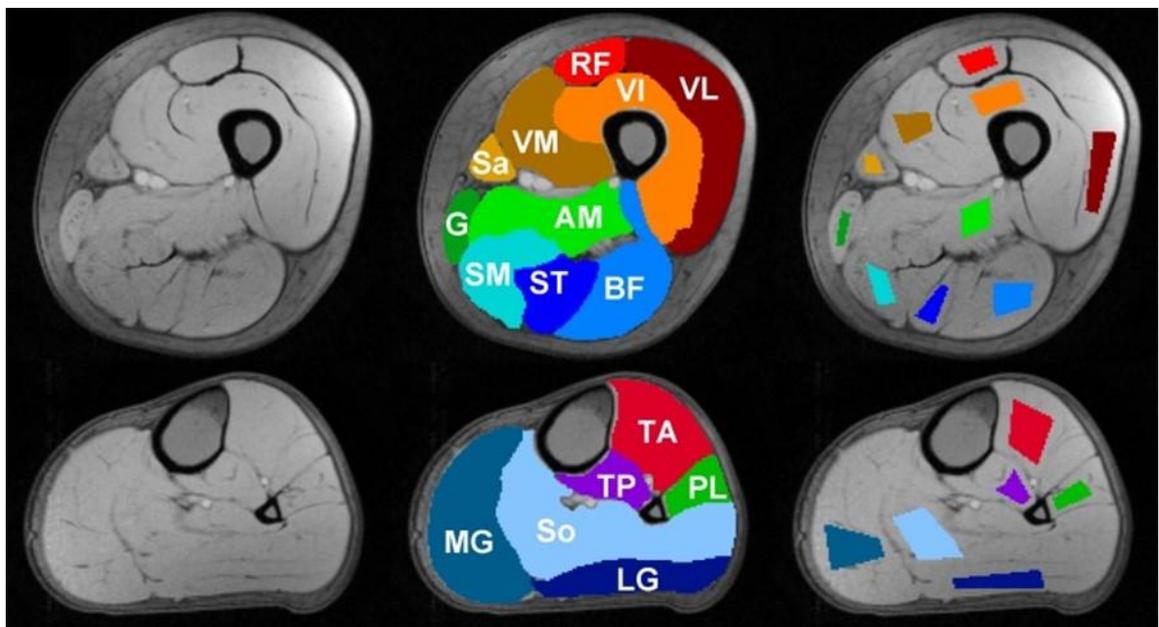
For the baseline scan the fifth most superior slice was used in the thigh and the sixth slice in the calf unless all the muscles were not visible, in which case an adjacent slice was selected. For the phase one participants who underwent test-retest imaging, ROIs for the second acquisition were drawn on the slice most similar in appearance to that used from the first acquisition. For phase 3, the ROI were drawn on the slice closest to that used for the phase 2 scan, determined on the basis of measured distance from bony landmarks (tibial plateau or tip of the fibular head) identified on the 3D-FLASH images. This was visually checked to confirm anatomical equivalence to the slice used

for the baseline scan by the radiologist, with agreement in every case. Slice selection based on a measured distance from a bony landmark has been shown to be superior to surface anatomy based localisation (Fischmann et al., 2014).

## 2.6.2 Region of interest definition

A radiologist with 4 years' post-specialist experience in neuromuscular imaging (Dr Fischmann, acknowledgements) defined "whole" and "small" regions of interest (ROI) for each phase 1 and phase 2 subject on the single slice defined above, at mid-thigh and mid-calf level of an unprocessed Dixon acquisition (TE=3.45ms) using ITK-SNAP (Yushkevich et al., 2006). "Whole" ROIs were defined to encompass the entire muscle cross-sectional area to the fascia whilst "small" ROIs were defined in a consistent anatomical location within each muscle to avoid contamination with fascia or vessels and to allow minor movement between acquisitions (Figure 2-5). Both types of ROI have been used by researchers previously (Hiba et al., 2012; Kan et al., 2009; Kim et al., 2010a; Sinclair et al., 2012a; Willis et al., 2013). Left and right limb ROIs were defined for the rectus femoris, vastus lateralis, vastus intermedius, vastus medialis, semimembranosus, semitendinosus, biceps femoris, adductor magnus, sartorius, gracilis, tibialis anterior, peroneus longus, lateral gastrocnemius, medial gastrocnemius, soleus and tibialis posterior muscles.

Figure 2-5: Region of interest definition



**Left:** Gradient echo (TE=345ms), one of the sequences used in three-point Dixon acquisition, of the left thigh (top) and calf (bottom). **Middle:** "Whole" ROI. **Right:** Small" ROI

**Thigh:** RF-rectus femoris, VM-vastus medialis, VI-vastus intermedius, VL-vastus lateralis, Sa-sartorius, G-gracilis, AM-adductor magnus, SM-semimembranosus, ST-semitendinosus, BF-biceps femoris (long head at this level)

**Calf:** *TA-tibialis anterior, EHL-extensor hallicus longus, EDL-extensor digitorum longus, P-peroneal muscle group; peroneus longus (PL)at this level), LG-lateral head of gastrocnemius, MG-medical head of gastrocnemius, So-soleus, TP-tibialis posterior, FHL-flexor hallicus longus, FDL-flexor digitorum longus*

To assess inter-observer reliability, the thesis author independently defined ROIs on one acquisition from each of the 15 phase one subjects with repeat examinations. ROI on the phase three scans were drawn by Dr Fischmann using the same procedure. This was done with direct reference to the ROI drawn on the phase two scans to ensure equivalent anatomical coverage.

### **2.6.3 Application of region of interest to maps**

The small ROIs were transferred to the co-registered parameter maps (FF, T1, T2, MTR). Minor adjustments to small ROI were made where imperfect registration meant ROI were no longer wholly within the target muscle. Whole muscle ROIs were transferred to the inherently co-registered FF maps only. The whole muscle ROIs from the unprocessed Dixon acquisition were not used for T2 and MTR analysis to avoid ROI contamination with non-muscle tissue, a particular problem for these measures, at the region boundaries due to minor subject movement between acquisitions. All generated maps were inspected visually for correct ROI placement and presence of artefact. ROI values originating from areas of gross artefact were excluded from the analysis.

## **2.7 Data analysis**

For each muscle custom written software (Dr Sinclair, acknowledgements) extracted mean T1, mean T2, mean FF and mean MTR from small ROI and mean FF and cross-sectional area (CSA) from whole ROI. After extraction all data were cross-checked for outliers and errors identified rectified. In addition to individual muscle values, summary measures for each parameter were created for relevant functional muscle groups (quadriceps, hamstrings, anterior tibial compartment and triceps surae) and for all muscles at thigh level and calf levels separately. For small ROI this was a simple mean, whilst for large ROI a weighted mean based on relative CSA was calculated. Total CSA was also calculated. Longitudinal changes were quantified on a muscle-by-muscle, parameter-by-parameter basis, and combined as detailed above to create separate all-muscle summary variables at thigh- and calf-levels.

To assess early pathological change in patients' muscles for which IFA lay within the normal healthy range, additional analyses included only muscles with mean FF less than the 95<sup>th</sup> percentile of the healthy control FF values for thigh and calf muscles separately. As a measure of the functional muscle CSA, the metric "remaining muscle

area”, defined as  $CSA \times (100 - FF)/100\%$ , was calculated, where CSA and FF refer to the whole muscle ROI values.

Longitudinal data were assessed by comparing baseline and follow up values, corrected for exact duration of follow-up on a muscle by muscle, parameter by parameter basis and combined as per the baseline data to create summary variables for thigh or calf level. Responsiveness of outcome measures was evaluated using the SRM, calculated as the ratio of mean change to standard deviation of change (Liang et al., 1990). SRM was categorised by magnitude according to Cohen’s suggestion: < 0.2 minimal responsiveness; 0.2-0.5 small responsiveness; 0.5-0.8 moderate responsiveness; >0.8 large responsiveness.

### **2.7.1 Data checking**

Overall the number of images excluded from the analysis was small: nine data-sets were missing or technically non-analysable: FF - thigh 1, calf 1; T<sub>1</sub> - thigh 2, calf 4; T<sub>2</sub> - none, MTR - calf 1. In the remaining data small fractions of individual ROIs were excluded due to local artefact, mostly B<sub>1</sub>-related signal drop-out: FF - thigh 1.7% (16/920), calf 2.4% (13/540); T<sub>1</sub> - thigh 24% (219/900), calf 12% (57/492), T<sub>2</sub> - thigh 5.4% (51/940), calf 0.2% (1/552), MTR - thigh 15% (142/920), calf 5.2% (28/540). In all subjects asymmetric B<sub>1</sub> deviations were observed with B<sub>1</sub> reduced anteriorly on the right and posteriorly on the left. This was evident at the calf level but more prominent in the thigh, particularly affecting the right rectus femoris and vastus medialis. This artefact prevented measurement of T<sub>1</sub> and MTR in these muscles for the majority of subjects. See also Table 4-5 and Table 4-6 which included the number of subjects analysed for each muscle and sequence.

### **2.7.2 Statistical analysis**

Phase one was analysed using SPSS 18 (SPSS inc., Chicago, Illinois), inter-muscle differences were assessed using ANOVA with post hoc comparisons using Bonferroni’s method. Inter-scan and inter-observer overall mean value differences were assessed using paired t-tests and reproducibility determined as mean absolute inter-scan and inter-observer differences, displayed on Bland-Altman plots with calculation of limits of agreement (Bland and Altman, 1986) and intra-class correlation coefficients (ICCs). Multivariate regression assessed the influence of demographic factors (age, gender, weight, height) on MRI measures: height showed no independent correlation with any MRI measure and was therefore excluded from the model. Pearson’s correlation coefficients between MRI measures were calculated.

Statistical analysis for phases two and three was performed using IBM-SPSS Statistics 20 (IBM Corp. Armonk, NY). The volunteer group of 29 subjects was arranged into two groups of 20 designated volunteer(IBM) and volunteer(CMT1A) (11 were controls to both groups). Summary statistics for each group of each quantitative parameter on a per muscle, and overall thigh or calf level were calculated (mean  $\pm$  standard deviation), using two-tailed t tests to assess for significant differences. Correlations were assessed using Pearson or Spearman correlation coefficients as appropriate. Due to the strong interdependence of fat fraction with the other quantitative parameters (Table 4-7), additional analysis was performed considering only muscles without significant fat infiltration, using a cut-off of 5% based on the volunteer data in this study. Multivariate linear regressions were performed to assess separately, in these muscles without significant IFA, the dependence of T2 and MTR upon disease versus control status while adjusting for the influence of residual FF as a covariate.

Baseline fat fractions were grouped into categories: <5%, 5-20%, 20-40%, 40-60% and >60% and rate of change in fat fraction compared according to these categories and the STIR categorisation. One-way ANOVA with post-hoc Tukey analysis was used to identify if any baseline MRI muscle characteristics were predictive of subsequent change.

## 3 Quantitative MRI in healthy volunteers

### 3.1 *Background*

It would be a mistake to rush to investigate quantitative MRI techniques in patients with neuromuscular diseases without assessing the methods' characteristics and performance in healthy volunteers. Whilst research involving patients is perhaps more immediately publishable and it is certainly true that ultimately the behaviour of outcome measures in patients is what matters most, characterising performance in healthy volunteers is a crucial step in rational outcome measure development.

With regard specifically to quantitative MRI of lower limb muscles, what may be gained from studies of healthy volunteers? First, at the simplest level a "normal range" can be established without which it is impossible to accurately judge if the results in patients are "abnormal". The terms are used in quotations as the reality is less straightforward than this. Second, potential experimental confounds can be established: Do values vary between muscles studied? Do values vary over the short term due to dietary intake or exercise for example? Do values vary between subjects depending on age, gender or size? All these are important questions since failure to consider the magnitude of these variations might reduce the responsiveness of an outcome measure, as over-large dependence on such factors could mask true differences related to a therapeutic intervention. Third and perhaps most importantly, the reliability of a technique can be examined. Within the more general class of health measurements, outcome measures are generally by definition longitudinal tools which means test-retest reliability (scan-rescan for MRI) is of the utmost importance. Again apparent variation between assessments due to measurement error could easily obscure what in many neuromuscular diseases might be a small true change between the study time-points.

It is beneficial to note at this early point the two fundamental categories of measurements possible with quantitative muscle MRI as they have different characteristics in healthy volunteers. Measurements of muscle **size** (width, circumference, cross-sectional area or volume) would naturally have large variation between muscles and subjects. Challenges to maximising their reliability mainly involve optimising repeatability of subject positioning, slice localisation and analysis methods. On the other hand, measurements of a muscle **tissue parameter** such as T2 or FF should have less inter- and intra-subject variation, certainly less than the 80,000 fold difference between the cross-sectional area of the smallest and largest muscles in the body (stapedius and gluteus maximus) (Voronov, 2003). Maintaining the same sequence parameters, and minimizing variation in instrumental performance,

is of greater importance than positional factor in the reliability of tissue parameter measurements.

### **3.1.1 Ideal outcome measure characteristics in healthy volunteers**

When considering fundamental characteristics of quantitative muscle MRI in healthy volunteers, it is helpful to review “ideal” outcome measure characteristics in this context, which summarised simply are minimal variability and maximal reliability (Kirshner and Guyatt, 1985). In order to maximise sensitivity, specificity, inter-scan reliability and responsiveness, an ideal outcome measurement would in healthy subjects have a narrow variance compared with the expected patient value ranges, and any physiological or inter-muscle or inter-subject variations subjects would be negligible, or at least of small magnitude and predictable.

Those current neuromuscular disease outcome measures which are expressible in terms of continuous variables stack up poorly against these ideals. Myometry assessments do not yield a narrow range of normal values, nor similar values between muscle groups in an individual subject. There is significant inter-subject variation according to gender and size, but also due to genetic factors and training level. Strength reduction with age is well known and documented both cross-sectionally and longitudinally (Andreassen et al., 2009), and may be a significant confound with respect to detecting longitudinal disease-related change. Neurophysiology-derived measurements, especially motor and sensory amplitudes, also show a wide normative range, with some of this variation dependent on demographic factors, and marked differences dependent on the specific muscle or nerve examined.

Reliability may also be assessed in healthy volunteer studies. Scan-rescan reliability is central to application of MRI as an outcome measure and should be maximised as far as practical. Whilst inter-observer reliability is of interest, in the case of MRI, unlike myometry and neurophysiology, this source of variability can be eliminated through central analysis by a single observer. Even in the event that a dataset was too large for analysis by a single observer, at least all time-points for each subject can be assessed by the same observer which, although it would be the aim for myometry and neurophysiology in multi-centre clinical trials, inevitably isn't possible on occasion. Furthermore, digitised images can easily be anonymised with respect to time-point (baseline or follow up) whereas myometry and neurophysiology collection and data extraction occur simultaneously, so blinding to time-point is not possible unless different observers are used, with resultant observer bias or inter-observer variability

Scan-rescan reliability in MRI measurements will be reduced by three categories of potential variation: physiological, acquisition and analysis. Physiological variation

refers to genuine changes in the tissue value, due to physiological factors which might vary between scans such as diet and exercise. Acquisition variation occurs during the scanning process, for example variation due to differences in patient positioning, sequence parameters or idiosyncratic artefacts. Analysis variation is largely the difference due to the subjective aspect of drawing regions of interest for analysis. Intra-observer variation can be directly assessed by repeated analysis of identical scans, whilst the first two sources of variation are more difficult to separate or measure directly. However, for use as an outcome measure it is their combined effect: scan-rescan reliability that is the key parameter, which should be maximised by reducing all three sources of variation as far as possible.

It is worth briefly distinguishing between random variation and systematic variation between scans, which in a longitudinal study may be a year or more apart. Random variation can be considered “noise” which might mask detection of true change between baseline and follow-up, resulting in a type II error (false negative), whereas systematic variation between time-points can create the false appearance of a difference when it does not exist, a type I error (false positive). Control groups and blinding are necessary components in the arsenal against this latter, potentially more serious error.

Healthy volunteer scanning is therefore essential in outcome measure development to establish “normal values” against which to compare patients, to identify potential demographic confounders and to determine the scan-rescan reliability of these measurements. This knowledge allows optimisation of study design, MRI protocol and analysis methods to maximise the potential of the outcome measures utilised. This chapter will review previous literature relating to quantitative MRI of skeletal muscle in healthy volunteers, describe results from the healthy volunteers imaged as part of this thesis, and finally discuss how these results impact on the application of quantitative MRI outcome measures in patient studies.

## **3.2 *Quantitative MRI variability and reliability in healthy volunteers***

This section will review published data on quantitative muscle MRI in healthy volunteers to provide context for the results which follow.

### **3.2.1 Published values and variation between muscles**

The concept of a normal or reference range is central to health related measurements. A result such as a serum potassium concentration has no diagnostic utility without the frame of reference of the usual range of values in the healthy population. This is

readily apparent when results are presented in unfamiliar units without a reference range. Reference ranges of diagnostic measurements are not absolute, they may differ depending on the age, gender or ethnicity of the subject; and alterations to the method of analysis may change the reference range entirely.

While still relevant, a reference range is not of primary importance for outcome measures where longitudinal change is of greater importance than the absolute values. Consider for example a weight loss intervention: the healthy population weight range is very wide, indeed the boundary between “normal” and “abnormal” weight is not defined based on the distribution of values in the general population, but on values associated with increased risk of disease. However, the measurement of weight change is simple, and ideal as an outcome measure in a clinical trial of a weight loss intervention. Similarly, different normative ranges have been reported for quantitative muscle tissue MR parameters in healthy volunteers, depending upon the exact scanner configuration, sequence and analysis used. This means that generally, a “reference range” defined in one study would be inappropriate to use to determine if a value obtained with different hardware, software or sequence parameters was abnormal. However relative differences, for example between muscles, should be consistent between studies. Outcomes measured as continuous variables are most commonly a relative comparison between baseline and end of study. This means for a longitudinal study, even if it is multi-site, scan-rescan reliability is more important than inter-scanner reliability.

### **3.2.1.1 T2 time**

The most commonly reported quantitative MR parameter of muscle is T2 relaxation time. Whilst exact sequence parameters vary between studies, the basic principle is the same. Images are acquired at multiple echo times and by plotting signal intensity against echo time, an estimate of the T2 relaxation time can be obtained. Values previously obtained in volunteers are listed in the table below. Reported healthy T2 of skeletal muscle in the calf (10 studies), thigh (5 studies), foot (one study) and upper limb (one study) lie in the range 25-45ms, with small intra-study standard deviations within a study of between 0.5-5 milliseconds (Table 3-1). It is fair to assume that the greater variation in values between studies owes more to methodological differences than true differences between the healthy volunteers. One early large study across eight sites suggested that field strength was not responsible for differences in T2 measurements (de Certaines et al., 1993), which is further supported comparing values obtained at 1.5T and 3T. There is one outlier in terms of T2 values, where the healthy volunteers had mean T2 values of 72ms (Maillard et al., 2004a). This suggests a difference in their methodology introducing a significant systematic difference in

measurements from all other studies, although the exact reason is not apparent in the paper. It does however emphasise that the message presented in their paper “A score of greater than 86 ms indicates inflammation” should be qualified with: ‘using our scanner, sequence and analysis method’ instead of implying an absolute cut off between normal and abnormal.

In all cases where a disease group was studied in addition to the volunteer group, the diseased group showed significantly higher values, suggesting that T2 measurement is effective at separating normal from abnormal skeletal muscle. Only one study with five controls analysed both thigh and calf muscles, and found T2 values higher in the thigh (Gloor et al., 2011) so there is insufficient data to know whether this is generalisable. Studies which measured T2 in more than one muscle found small but significant differences between muscles as outlined in Table 3-2 (Gloor et al., 2011; Hatakenaka et al., 2001; Psatha et al., 2012; Schwenzer et al., 2009). Compared with soleus, which is the most commonly analysed muscle, the reported T2 time in tibialis anterior is shorter. Either increased intramuscular fat content or higher Type 1 fibre concentration is the given explanation. These differences were small compared with values obtained in pathological muscle.

Table 3-1: Reported T2 in skeletal muscle of healthy volunteers

Region	Reference	MRI	Echo	N	Muscle	Side	ROI	T2 value	Patient
Calf	(Ababneh et al., 2008)	3T	16	6	TA	Right	Small	32 ± 1.6ms	Exercise
Calf	(Bryan et al., 1998)	NS	8	8	TA	Both	NS	33.4 ± 1.5ms	ALS
Calf	(Hatakenaka et al., 2001)	1.5T	4	59	So, GM	Right	Small	26.9-30.3 ± 0.6-1.8ms	No
Calf	(Hatakenaka et al., 2006)	1.5T	NS	11	So, GM	Right	Small	33.5 ± 1.7 30.7 ± 1.1	No
Calf	(Hiba et al., 2012)	1.5T	2	6	TA	Both	Whole	42.4 ± 2.9ms	Myotonic dystrophy
Calf	(Huang et al., 1994)	1.5T	NS	2	TA	Right	Medium	25-27ms	Duchenne
Calf	(Phoenix et al., 1996)	1.5T	4	3	TA, So, GM	Both	Full	28.5ms	Muscular Dystrophy
Calf	(Ploutz-Snyder et al., 1997)	1.5	64	9	TA	Right	Small	29.3 ± 0.7ms	Exercise
Calf	(Psatha et al., 2012)	3T	8	18	TA So GM	Single	Medium	27 ± 2.5ms 34.6 ± 2.9ms 34.4 ± 2.9ms	Immobilisation
Calf	(Sakurai et al., 2011)	1.5T	NS	15	TA So	Single	NS	29.8 ± 1.6ms 32.5 ± 1.3ms	No
Calf	(Schwenzer et al., 2009)	3T	32	23	TA, So, GM, GL	Right	Medium	32-35 ± 1-3ms 34-40 ± 2-6ms	Young Old
Calf Thigh	(Gloor et al., 2011)	1.5T	32	5	10 calf 11 thigh	Both	Whole	38±1 ms calf 33±1 ms thigh	OPMD
Thigh	(de Certaines et al., 1993)	0.1-1.5T	Var	100	8 thigh	Right	Medium	32-47 ± 1-4ms	No
Thigh	(Maillard et al., 2004a)	1.5T	16	20	Ant/Post/ Med comp	Left	Medium	78 ± 2ms	Juvenile DM
Foot	(Bus et al., 2002)	3T	11	8	Whole CS	Left	Whole	40-50ms	Diabetes
Upper limb	(Ploutz-Snyder et al., 2006)	1.5	64	6	Biceps Triceps	Single	Full	29.9 ± 0.5ms 30.4 ± 0.7	Stroke

NS: Not specified; TA: tibialis anterior; So: Soleus; GM: Gastrocnemius medial head  
GL: Gastrocnemius lateral head ROI: Region of interest type; Var: variable; CS: Cross-section

Table 3-2: T2 times in different calf muscles

Reference	So	TA	MG	LG
(Schwenzer et al., 2009)	34 ± 1ms	32 ± 1ms	35 ± 3ms	35 ± 2ms
(Hatakenaka et al., 2001)	29.4 ± 1.3ms	X	27.2 ± 1.2ms	X
(Sakurai et al., 2011)	32.5 ± 1.3ms	29.8 ± 1.6ms	X	X
(Hatakenaka et al., 2006)	33.5 ± 1.7ms	X	30.7 ± 1.1ms	X
(Psatha et al., 2012)	34.6 ± 2.9ms	27 ± 2.5ms	34.4 ± 2.9ms	X
(Gloor et al., 2011)	42 ± 3ms	33 ± 1ms	43 ± 6ms	40 ± 5ms

TA: tibialis anterior; So: Soleus; MG: Medial head of gastrocnemius; LG: Lateral head of gastrocnemius; X: not reported

### 3.2.1.2 T1 time

Within the quantitative MRI literature, the T1 relaxation time of healthy skeletal muscle is much less commonly reported. T1 relaxation time is dependent on field strength, with theoretical models available to predict this, and demonstrated in an early multicentre trial (Henriksen et al., 1993). There is insufficient data to make any assessment of differences between muscles. Therefore, T1 values cannot be directly compared between studies, and multi-centre studies will need to ensure equivalent sequence parameters at the different sites with an appropriate inter-site standardisation process. Careful corrections would need to be applied if scanners with different field strengths were used at different sites.

Table 3-3: Reported T1 values in skeletal muscle of healthy volunteers

Region	Reference	MRI	N	Muscle	Side	ROI	T1 value	Patient
Calf	(Bryan et al., 1998)	NS	8	TA	Both	NS	861 ± 21 ms	ALS
Calf	(Sakurai et al., 2011)	1.5T	15	TA So	NS	NS	1219 ± 223 1167 ± 203	No
Calf	(Huang et al., 1994)	1.5T	2	TA	Right	Medium	1300 ± 100	Duchenne
Thigh	(de Certaines et al., 1993)	0.1 - 1.5T	100	8 thigh	Right	Medium	202 ± 26 ms 1109 ± 126 ms	No

NS: not specified; TA: tibialis anterior; So: soleus, ROI: region of interest type

### 3.2.1.3 Fat fraction

Intramuscular fat accumulation (IFA) is a feature of chronic muscle damage common to both neurogenic and myopathic diseases. Quantification of IFA therefore represents a promising biomarker of disease progression in chronic neuromuscular diseases. There are a number of publications quantifying fat content in skeletal muscle of healthy volunteers (Table 3-4), however direct comparisons are limited by different MR

methods (MRI vs MR spectroscopy), as well as differences in sequence and analysis method.

### **3.2.1.3.1 Methods of fat fraction estimation**

The simplest method to estimate fat infiltration within a muscle is segmentation based on signal intensity on a T1w image. This method has been used in both patients (Ravaglia et al., 2010) and volunteers (Holmbäck et al., 2002). However, these sequences are prone to systematic artefact and the classification of each voxel is binary (muscle or fat), so this technique is less commonly used. Another technique utilises T2 modelling to estimate tissue fat fraction (Gloor et al., 2011; Kan et al., 2009). An alternative method utilises sequence parameters to create an image where signal is primarily derived from protons in fat, then normalises to an internal reference with known fat concentration such as bone marrow. However the most commonly used methods rely on differences in resonance frequencies between protons in fat and water caused by their “chemical shift” differences (Reeder et al., 2012). Chemical-shift can be exploited to separate fat and water signals using either MR Spectroscopy (MRS) or conventional MR imaging sequences. Single voxel (SV) MRS methods examine the spectra of protons (<sup>1</sup>H) within a single voxel positioned within a muscle with reference to an anatomical reference image, from which the fat and water peak areas are compared to calculate a fat fraction for that volume of tissue. If methods with adequate spectral resolution are used, intramyocellular and extramyocellular lipids can be recognised as separate spectral peaks and quantified separately. Multi-voxel techniques apply the same methods across multiple voxels in the same acquisition to provide a degree of spatial resolution, although this greatly increases scanning time.

The three-point Dixon fat-water separation method (Glover and Schneider, 1991) and other related MR imaging techniques take advantage of the chemical shift to design an acquisition scheme yielding “fat only” and “water only” images. These two images are combined according to the formula:

$$\text{Signal Fat Fraction} = \text{Fat Signal} / (\text{Fat Signal} + \text{Water Signal})$$

#### *Equation 3-1: Fat fraction*

This may be expressed as a decimal or percentage fraction, with the latter notation used in this thesis. This provides a fat fraction map with high spatial resolution which can then be analysed to determine the mean fat fraction within any desired region of interest. If all possible sources of systematic error are reduced or eliminated in this process, then the measurements of signal fat fraction will correspond to visible proton density fat fraction. This is defined as the ratio of the concentration of mobile protons from fat (triglycerides) and the total concentration of protons from mobile triglycerides and mobile water. The PDFF is a fundamental property of tissue and reflects the

concentration of fat within that tissue. It has been proposed as the best MR biomarker of fat infiltration (Reeder et al., 2012).

The technical aspects of the different methods of quantification of fat infiltration are beyond the scope of this discussion. However, it is important to note that widely different methods of acquisition and analysis exist in the literature and the units of measurement also vary. Choice of method largely depends on the purpose of the study. A time intensive MRS method which can distinguish intramyocellular and extramyocellular lipid components is important in metabolic (Shen et al., 2008) or sports (Tamura et al., 2008) research, but for neuromuscular disease biomarkers spatial resolution, reliability and acquisition time are more important considerations. Therefore MRI based methods such as Dixon are generally preferred in this context. As with T2 estimations, the fat quantification methods used need to be carefully taken into account when comparing studies.

### ***3.2.1.3.2 Reported values, variation between muscles***

Within the limitations noted above, reported values for fat fraction in skeletal muscle of healthy volunteers are summarised below. Where possible values are converted to a “fat fraction percentage” to be more readily comparable. Two studies comparing two different fat fraction estimation techniques (Gloor et al., 2011; Schick et al., 2002) are of particular interest.

It is interesting to note differences in three studies which report fat fractions in tibialis anterior muscle using the Dixon fat-water separation method. Psatha and colleagues (2012) found a mean fat fraction in tibialis anterior of 3% using ROI drawn within the muscle avoiding fascia and vessels. Gloor and colleagues (2011) had a mean value of 5% using ROI drawn encompassing the whole cross section at a single slice, whilst Hiba and colleagues (2012) reported a median value of 11% when ROI were used to encompass the whole volume of muscle. The differences may be due to the increasing amount intramuscular and extra-muscular fat included within ROI, although the application of these types of ROI on the same dataset would clarify this further.

One focus of <sup>1</sup>H-MR spectroscopy has been quantification of intramyocellular lipid, to investigate its role in metabolism, training and metabolic diseases such as diabetes. Intramyocellular lipid content varies between muscles, with higher concentrations in muscles with the highest proportion of Type 1 fibres, trained endurance athletes, obese individuals and if a high lipid diet is consumed. However, the total amount of IMCL is small: 0.24% by volume on average (Howald et al., 2002), so even in muscles with high Type 1 content in trained athletes the concentration is less than 1% by volume. This amount is negligible compared with the total fat content reported in patients with

neuromuscular diseases, so is likely to be of less importance in this setting. Some investigators provide estimates of total lipid content using MRS – although for small voxel sizes the EMCL varies depending on exact position (Shen et al., 2008) so for this purpose large voxels or a multi-voxel MRS technique are preferable. Values of total fat fraction with MRS are very similar to those obtained using chemical-shift MRI methods.

Table 3-4: Reported fat fraction in skeletal muscle of healthy volunteers

Reference	MRI	Method	N	Tissue	Fat value	Comment
<b>CHEMICAL SHIFT MRI METHODS</b>						
(Psatha et al., 2012)	3.0T	3pt Dixon	18	TA, So, MG (uncast leg) Moderate ROI	3% 5% 5%	In cast other leg day 5 but stable
(Hiba et al., 2012)	1.5T	3pt Dixon	6	TA: both. Full muscle volume	11 ± 2%	Converted from ratio Myotonic dystrophy
(Gloor et al., 2011)	1.5T	2 pt Dixon And Steady State Free Precession	5	Thigh and Calf muscles Full ROI	2pt Dix: 8.3% ±1.8% thigh 6.2 ±0.8% calf SSFP 4.1 ± 1.1% thigh 2.6±0.6% calf	OPMD
(Hatakenaka et al., 2006)	1.5T	2-pt Dixon method	11F	Right So, MG. Small ROI	Fat deposition ratio 0.915±0.014 0.904±0.005	Also rabbits
<b>1H-MRS METHODS</b>						
(Torriani et al., 2012)	3T	SV-1H-MRS Lipid fraction	8	Right TA and So Small ROI	2.5 ± 2.0% 7.4 ± 4.0%	Duchenne
(Ortiz-Nieto et al., 2010)	1.5T	H1-MRS using bone marrow as internal reference	20	Right Thigh Right Calf	2.7/3.2±1.0/1.3% 1.2/1.3±0.4/0.4% (M/F)	Also IMCL. No difference Male:Female
(Schick et al., 2002)	1.5T	MRS and Fat selective imaging (bone marrow ref.)	30	Right TA, So	2.0/3.1±0.7/1.2% 3.1/3.3±1.1/1.5% (M/F)	Also examined correlation spectroscopic and imaging techniques
(Pfirrmann et al., 2004)	1.5	SV-1H-MRS	30	Supraspinatus	Apparent lipid content (FF) 13.7%	Supraspinatus tears
(Shen et al., 2008)	1.5T	SV-1H-MRS IMCL/Water	5 5	Right TA	0.040 ± 0.010 0.032 ± 0.002	Obese Lean Also Multivoxel
(Howald et al., 2002)	1.5T	SV-1H-MRS IMCL (%ww)	10	Right TA	0.24 ml/100mLww	Compared to biopsy
(Sakurai et al., 2011)	1.5T	MRS for IMCL	15	TA So	2.3 ± 1.3 7.7 ± 3.5 IMCL/Cr	Side/ROI type not given
(Forbes et al., 2014)	3T	SV-1H-MRS	31	VL Sol	0.027 ± 0.017 0.015 ± 0.006	5-14 year old boys Duchenne
(Tamura et al., 2008)	?	SV-1H-MRS	14	Right TA Right So	IMCL (Fat-Cr ratio): 1.15/1.81 TA 5.1/5.92 So	Sprinters/ Endurance athletes. Also different diets
<b>OTHER METHODS</b>						
(Schwenzer et al., 2009)	3T	Spoiled gradient-echo sequence with spatial-spectral excitation	11 +	Right TA, So, MG, LG	1.2-3.5 ± 0.4-1.1% +	Young and old
(Kan et al., 2009)	3T	Multispin echo	7	Right Calf muscles	2±2% TP 9 ±5% MG	FSHD

TA: tibialis anterior; So: soleus; MG: medial head of gastrocnemius; LG: lateral head of gastrocnemius; ROI: region of interest; SV: single voxel; IMCL: intramyocellular lipid; Cr: creatine

### 3.2.1.4 MTR

Measurements derived from the magnetisation transfer properties of skeletal muscle are less common. From those listed in the table below it is clear that healthy muscle displays significant magnetisation transfer effect but as expected the value obtained is dependent on both field strength and sequence implementation. With the limited data available there are not large differences between values of muscles in the lower leg, and no assessments have been published of thigh values in healthy volunteers.

Table 3-5: Reported MTR in the skeletal muscle of healthy volunteers

Reference	MRI	Method	N	Tissue	MTR	Comment
(Schwenzer et al., 2009)	3T	TR 36ms, TE 5.2ms, MT prepulse 1200 Hz duration 9.472 ms, flip angle MT pulse 750°	11 + 12	Right TA, Sol, MG, LG Moderate ROI	43-47 ± 3-4 p.u.	Young and old Difference in TA only but possible B1 effects
(McDaniel et al., 1999)	0.5T	TR 600ms, TE 30ms, MT prepulse 200 Hz duration 16 ms	10	Right calf 5 x muscles Small ROI	38-41 ± 1-3 p.u.	Also LGMD No significant correlation with age
(Sinclair et al., 2010)	3T	Full qMT	10	Right Calf 7 x muscles Small ROI	27-36 ± 2-4 p.u.	Dedicated qMT protocol. Variation due to uncorrected B1 inhomogeneity
(Sinclair et al., 2012a)	1.5T	TR 1500ms, TE 14ms, flip angle 70°.	10	Both calves, 4x compartments	51 ± 2 p.u.	CIDP (39 p.u.), CMT (42 p.u.)

TA: tibialis anterior; So: soleus; MG: medial head of gastrocnemius; LG: lateral head of gastrocnemius; ROI: region of interest; p.u.: percentage units; qMT: quantitative magnetisation transfer; LGMD: limb girdle muscular dystrophy; CIDP: chronic inflammatory demyelinating polyneuropathy; TR: repetition time; TE: echo time

### 3.2.2 Physiological variation

Physiological variation, if not characterized and controlled for is a major confounder in health measurements used both for diagnosis and as outcome measures. For example, serum cortisol shows diurnal variation so timing of the sample is required for interpretation, and growth hormone levels show such wide variation that a random level is of little value and suppression or stimulation tests are required. For skeletal muscle quantitative MRI measurements there are two major potential sources of physiological

variation within an individual: diet and exercise. Both of these may be considered in the short term: physical activity or food intake in the 24 hours preceding the MRI; or considered in the medium term: effects of training or general diet. These factors have been reported to some extent in the literature to date.

### **3.2.2.1 Effect of exercise and training**

The effect of exercise on muscle T2 in volunteers has been examined in a number of studies. Two studies have measured the T2 in tibialis anterior before and after exercising this muscle to fatigue within the scanner. These showed mean increases in T2 of 11ms (Ababneh et al., 2008) and 9ms (Ploutz-Snyder et al., 1997) with subsequent decay toward baseline with a half-life of 8 minutes (Ababneh et al., 2008). Other studies have seen a ~3ms increase in upper arm muscles (Ploutz-Snyder et al., 2006) and 2ms increase in thigh muscles (Maillard et al., 2005) following exercise, with this returning to baseline when repeated at 30 minutes in the latter study. Taken together these studies do demonstrate an immediate effect of exercise on muscle T2, perhaps relating to changes in blood volume (Ababneh et al., 2008), however the effect lasts less than 30 minutes.

Medium-term changes in muscle use, whether training or immobility, have also been investigated, though mostly with regard to muscle size, which not unexpectedly increases with training (Nakai et al., 2008) and reduces with immobility (Miokovic et al., 2012). One study, in addition to atrophy, quantified T2 in lower leg muscles following immobilisation due to ankle fracture. They found greatest increase in T2 time in soleus muscle with a 10ms increase, with lesser increases in gastrocnemius and tibialis anterior at 5ms and 3ms respectively. This effect seemed to develop after 15 days of complete immobilisation so would be less relevant in clinical trials in neuromuscular diseases unless joint arthrodesis was performed during the study period.

### **3.2.2.2 Effect of diet**

The short term effect of diet on intramyocellular lipid content has been examined in athletes (Tamura et al., 2008). Under extremes of diet (10% versus 60% calories from fat), differences in the intramyocellular lipid content in calf muscles can be detected in endurance athletes but not sprinters. However intramyocellular lipid content is a small fraction of total muscle volume (0.24%) (Howald et al., 2002), and these differences occurred with combined training and diet extremes, so are unlikely to result in significant day-to-day variation when assessing total fat fraction (combined intra- and extracellular lipid content) in subjects with more typical lifestyles. Long-term changes in diet will result in change in weight and body fat distribution, which may result in significant changes in extramyocellular lipid content, considered in 3.2.3.3 below.

### 3.2.3 Inter-subject variation

One of the major limitations of current outcome measures in neuromuscular disease, such as strength, is the wide range of values which can be normal. For example, knee extension strength varies hugely between a 20-year-old male professional rugby player and a healthy 80-year-old woman. What follows is that the rugby player could lose 75% of their strength in this muscle but still have “normal strength” if an age/gender specific reference range is not used, and even then a similar difficulty would apply with comparison of the 20-year-old professional rugby player to a sedentary 20-year-old. Proper interpretation of health measurements and optimal application of outcome measures is greatly aided by better understanding of these inter-subject differences. Furthermore, changes with increasing age will cause confounding in longitudinal studies.

#### 3.2.3.1 Effect of age on quantitative MRI parameters

Changes with age can cause confounding in longitudinal studies of disease progression. Whilst gender almost certainly stays constant through a longitudinal study, age certainly does not and so will influence outcome measure responsiveness. A recent publication in CMT suggested that the apparent progression observed in adult patients was not different from the effects of ageing in matched healthy volunteers (Verhamme et al., 2009). This finding is at odds with clinical observations, and an alternative explanation is that the markers of progression used in this study, neurophysiology and muscle strength testing, were insufficiently sensitive to show change in the time-frame of the study, especially with changes of healthy ageing confounding measurement.

As might be expected, total muscle mass is negatively correlated with age, and this is reflected in both cross-sectional and longitudinal studies involving **size** measurement of lower limb muscles. Andreassen and colleagues (Andreassen et al., 2009) performed lower leg and foot imaging at a 9-12 year interval and found a loss of calf muscle volume of 1.7% per annum, although minimal (0.2%/annum) change in foot muscle volume. In a cross-sectional MRI study (Jubrias et al., 1997) a 1.4% per year decline in quadriceps cross-sectional area was seen in subjects aged between 65 and 80 years. This result is consistent with a longitudinal CT study with 1678 participants aged 70-79 at baseline, where a 0.8% per annum reduction in muscle cross-sectional area at mid-thigh level was observed (Delmonico et al., 2009). Interestingly, in both of these studies the reduction in muscle strength per annum was twice that of cross-sectional area, indicating that reduction in force with age is not solely due to reduced muscle size, but also associated with a reduction in the specific force (maximum force

per unit area) of the muscle. All these studies make clear that if a muscle **size**-based measurement is utilised as an outcome measure, decline with healthy ageing will be a potentially confounding factor similar to when neurophysiology or myometry are used.

There are also some data available with regard to **tissue parameter** muscle measurements with age, though to date these are derived from cross-sectional analyses. Schwenger and colleagues (Schwenger et al., 2009) have published data on quantitative MRI measurement of T2, Dixon fat fraction and MTR of the muscles in the right calf at 3T. They compared a group of 12 younger adults, mean age 31, with 11 older adults, mean age 66. Utilising subtotal ROI of tibialis anterior, soleus, medial and lateral gastrocnemius muscles they showed T2 and fat fraction differed between the groups in all muscles. The absolute difference however was small: with T2 in the younger group 34ms versus 37ms in the older group. This corresponds to a less than 0.1ms change in T2 per year of increasing age, although caution is needed when using cross-sectional data to predict longitudinal change. Similarly, the mean fat fraction in the younger group was 2.8% versus 6.6% in the older group, corresponding to approximately 0.1% change per year. For the MTR measurements a significant difference was only seen in tibialis anterior (47p.u. in the younger group, 44p.u. in the older group) but not in other muscles. A strong relationship between T2 and fat fraction was observed, leading the authors to conclude that the changes in T2 resulted from increased extramyocellular lipid content with age.

Another study quantified T2 relaxation time in gastrocnemius and soleus muscles of 59 volunteers aged 22 to 76 (Hatakenaka et al., 2001). They showed a positive correlation between age and T2 in gastrocnemius but not soleus, which together with animal studies led them to conclude that the difference was due to age-related Type II fibre atrophy. Again the absolute difference in T2 was small, increasing from 26 to 29ms in the age range examined – much less than 0.1ms per annum. Although caution is needed in interpreting cross-sectional data as a surrogate for longitudinal data, it is clear that changes in measured tissue parameters are likely to show minimal change over the likely duration of a clinical trial (1-5 years) compared with changes which might be seen in muscle size measurements.

### **3.2.3.2 Effect of gender on quantitative MRI parameters**

As would be expected from muscle strength differences between the sexes, there are differences in muscle **size** measurements between males and females (Jubrias et al., 1997). Indeed in children, differences in strength observed between genders was fully explained by differences in muscle cross-sectional area (Wood et al., 2004). The only study specifically examining whether **tissue parameter** measurements show gender

related bias did not show any gender difference in IMCL or EMCL of thigh and calf muscles (Ortiz-Nieto et al., 2010). More information in this area is clearly needed.

### **3.2.3.3 Effect of body habitus on MRI parameters**

It would be no surprise if larger people have larger muscles, though in fact this hasn't been systematically examined. It might also be hypothesised that obese individuals would have greater proportions of intramuscular fat and this has been examined in the metabolic literature, unsurprisingly given the global importance of obesity, diabetes and related conditions. Perhaps surprisingly therefore, investigators have found no correlation between muscle fat and BMI for subjects with BMI less than 30 (kg/m<sup>2</sup>), although three subjects in the study with BMI >30 had increased muscle lipid content (Schick et al., 2002). Similarly, there was no correlation between thickness of subcutaneous fat and lipid content in the muscle. Other studies have examined more specifically the metabolically active intramyocellular lipid content and found higher IMCL levels in obese subjects (Shen et al., 2008), although as noted previously this is also true for highly trained endurance athletes and the absolute amount this contributes per weight muscle volume is small (Howald et al., 2002). Overall therefore there is evidence which suggests that, similar to age and gender, any effect of body size of muscle tissue parameter measurements is small; however a more systematic assessment of this is warranted.

### **3.2.4 Reliability**

As already discussed test-retest reliability is absolutely critical to outcome measure responsiveness: it determines the "noise" and as disease progression in many neuromuscular diseases is slow, can easily mask the "signal" we are trying to detect. For example if a measurement such as muscle strength testing shows variation of 10% or repeat testing the next day, it is very difficult to demonstrate a 2% annual change in strength without either a very large cohort of patients or a long follow up period. Optimising reliability is critical to maximising the responsiveness of any outcome measure, so the relative paucity of true scan-rescan reliability assessment in the quantitative muscle MRI literature needs redressing.

For **size** measurements such as muscle area or volume, intra-rater (Commean et al., 2011; Delmonico et al., 2009; Fortin and Battié, 2012; Gille et al., 2011; Holmbäck et al., 2002; Hoyte et al., 2009, 2011; Hu et al., 2011; Humbert et al., 2008; Ranson et al., 2006; Smeulders et al., 2010; Springer et al., 2011; Tingart et al., 2003; Wood et al., 2004) and inter-rater (Commean et al., 2011; Gille et al., 2011; Hiba et al., 2012; Holmbäck et al., 2002; Hoyte et al., 2009, 2011; Hu et al., 2011; Kilgour et al., 2012; Smeulders et al., 2010; Springer et al., 2011; Tingart et al., 2003) reliability has been

reported and is generally excellent. As would be expected intra-rater is higher than inter-rater reliability, although the latter source of variation can be eliminated through the trial design ensuring all scans from one individual are assessed by one observer. Inter-scan reliability has been assessed for cross-sectional area of neck muscles (Kilgour et al., 2012), which was high when reported as intra-class correlation coefficients (0.95-0.97) but the coefficients of variation were high, between 22 and 26%. Sproule and colleagues assessed thigh muscle volume in 8 patients with spinal muscular atrophy, with immediate rescanning finding extremely high ICC (>0.99), though without reporting limits of agreement or coefficients of variation to allow analysis of what interval change could be considered significant. Inter-scan reliability of forearm muscle volume estimation has also been reported to be high with ICC >0.99, though this figure hides significant inter-scan variation such that the smallest detectable change in muscle volume was reported to be 7% (Smeulders et al., 2010). Finally Clark and colleagues (Clark et al., 2007) performed a useful assessment of test-retest reliability over 4 weeks of a range of assessments of ankle plantarflexion function including muscle cross-sectional area of soleus, medial gastrocnemius and lateral gastrocnemius. Again ICC were relatively high (0.96-0.99) but limits of agreement were variable 37%, 6% and 14% respectively. These muscles vary in cross-sectional area relatively rapidly along their length, so inconsistent slice positioning is the most likely cause in this case. All these examples show that while inter-scan reliability of size measurements is apparently high, this is due to the high inter-subject variation (see Equation 1-1) and limits of agreement which describe the range where apparent differences could be due to measurement error are in fact wide.

Fewer studies have examined reliability for measurements of muscle tissue parameters. One study examined Inter-observer and intra-observer reliability of mean T2 measurements within calf muscles in a mixed group of patients and volunteers, with variance of 1.1ms and 0.6ms respectively (Huang et al., 1994). Another study showed fat-to-water ratio intra-observer variance of 2.5% in a group of patients with myotonic dystrophy, with high reliability in more severely affected patients. Reliability was better than for residual volume measurements (5.2%) and myometry measurements (11.2%) (Hiba et al., 2012). Scan-rescan reliability has been examined only for proton spectroscopy methods of lipid estimations: for total lipid content in supraspinatus muscle (Pfirrmann et al., 2004), and intramyocellular lipid content of tibialis anterior muscle (Shen et al., 2008). For the remainder of quantitative MRI methods, inter-scan reliability has not been systematically reported, despite its key role in outcome measure performance.

### **3.2.5 Summary**

Published data of quantitative muscle MRI in healthy volunteers to date has established that for **size** measurements there is significant inter-subject variation which relates to gender and age, and that longitudinal change in muscle size can be quantified with training and immobility. This high inter-subject variation results in high reliability of measurements when assessed by intra-class correlation coefficients, but rather less reliability when assessed by limits of agreement, which are more relevant indices of reliability for longitudinal outcome measure assessments. On the other hand, tissue MRI parameters such as T1, T2 and MTR vary more between studies of healthy volunteers than between participants within studies, demonstrating that absolute “normal ranges” for these measurements are inappropriate. Although data are sparse, it appears inter-subject variation in these measurements is small with little or no dependence on age/gender/weight. Small differences between muscles and good reliability are suggested by small studies, but more rigorous assessment of this is needed.

### **3.3 *Healthy volunteer phase aims***

The specific aims of phase 1 of this thesis, the healthy volunteer phase are to:

1. Determine the normal range of quantitative muscle MRI parameters (size, T1, T2, MTR, fat fraction) with our methods (scanner, coil, positioning, sequences, post-processing and analysis)
2. Assess between-muscle differences in quantitative muscle MRI parameters to demonstrate the sensitivity of our methods to small physiological variation
3. Determine the inter-subject variation of quantitative muscle MRI parameters and the determinants of this variation
4. Assess the reliability of these measurements: inter rater and scan-rescan
5. Assess the relationship between the quantitative muscle MRI parameters and also the relationship between quantitative measurements and qualitative MRI assessments.

## 3.4 *Results*

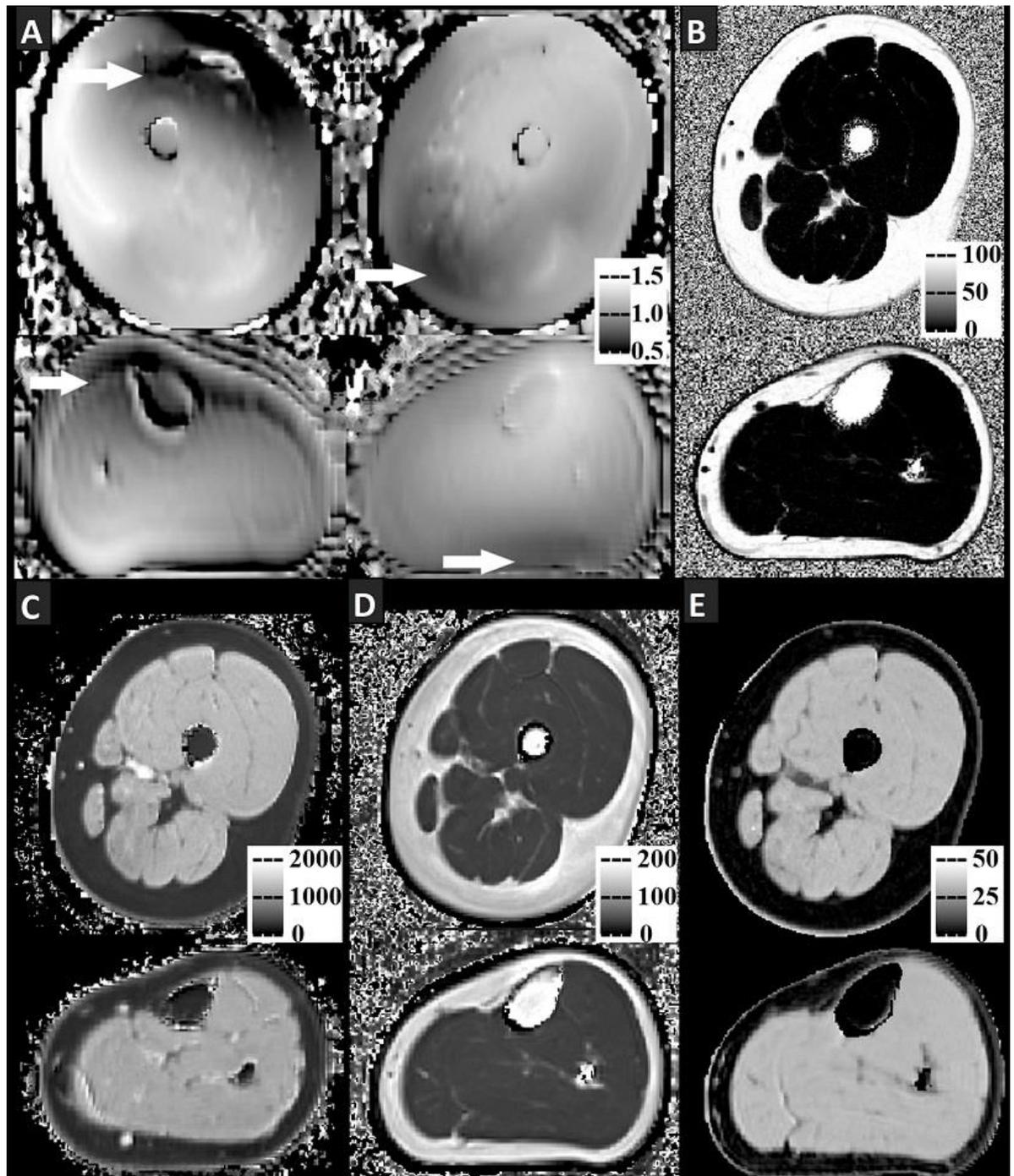
### 3.4.1 **Subjects and scan schedule**

With local research ethics committee approval and written consent 47 healthy volunteers were studied: 23 males, age  $44.4 \pm 17.0$ , 21.5-81.0y; height  $171 \pm 9$ , 150-188cm; weight  $73 \pm 16$ , 44-115kg; body mass index  $25 \pm 4.7$ , 17-41 kg/m<sup>2</sup> (mean  $\pm$  SD, range). Data presented includes 19 healthy volunteers scanned during phase 1 and 28 healthy volunteers during phase 2 (see 2.2). Fifteen healthy volunteers from phase 1 underwent repeat imaging after approximately 2 weeks with identical imaging parameters.

### 3.4.2 **General appearance**

Sample parameter maps are shown in Figure 3-1. Example ROI as drawn on one of the unprocessed Dixon acquisitions are shown in Figure 2-5. Significant signal inhomogeneity is notable in this – with lesser signal intensity anteriorly on the right and posteriorly on the left for both thigh and calf, though the effect was greater for thigh than calf. This was consistent amongst all subjects, and the same pattern was seen in the B1 field map (Figure 3-1-B). The nature of the three-point Dixon method and T2 mapping where the maps comprise a ratio of signal intensities or an estimated decay constant (see 1.4) meant the generated FF (Figure 3-1-C) and T2 (Figure 3-1-E) maps visually appeared free of this inhomogeneity, although occasionally artefact was present in the antero-medial right thigh when there was complete signal drop-out. Both MTR and T1 maps were generating using a method with a correction based on the co-registered B1 field map (see 2.4.4.2 and 2.4.4.5) and the resulting maps (Figure 3-1-D and Figure 3-1-F) visually appeared free of field inhomogeneity although for both these sequences signal drop-out antero-medially in the thigh made this area non-analysable in the majority of subjects. Specifically, artefact was present within right rectus femoris in 45/47 T1 maps, 41/47 MTR maps and within right vastus medialis in 35/47 T1 maps and 33/47 MTR maps.

Figure 3-1: Sample images from a single healthy volunteer (24 year old male)



A:  $B_1$  field map of both thighs and both calves demonstrating reduced  $B_1$  anteriorly on right and posteriorly on left (arrows). B: Fat fraction map left thigh and calf in %. C:  $T_1$  map in ms at left thigh and calf level. D:  $T_2$  map left thigh and calf in ms. E: MTR map left thigh and calf in p.u. All images are axial with standard orientation (anterior at top of image, right hand side at left of image).

### 3.4.3 Individual muscle values

Individual muscle values from small ROI for each MRI measure in all 47 subjects are shown in Figure 3-2. FF were similar in the same muscles in left and right limbs, confirming  $B_1$  inhomogeneity did not unduly influence these measures. Between muscles, FF differed significantly (ANOVA,  $p < 0.001$  at both calf- and thigh-level). Group-mean sartorius FF was higher than all other thigh-level muscles ( $p < 0.01$  for semimembranosus,  $p < 0.001$  for all other muscles) whilst the rectus femoris FF was lower than most other thigh muscles ( $p < 0.01$  vs. gracilis, vastus lateralis;  $p < 0.001$  vs. sartorius, semimembranosus, biceps femoris and adductor magnus). Similarly in the calf, soleus FF was highest ( $p < 0.05$  vs. peroneal,  $p < 0.01$  vs. medial gastrocnemius,  $p < 0.001$  vs. each remaining muscle) whilst tibialis anterior FF was the smallest ( $p < 0.01$  vs. medial and lateral gastrocnemius,  $p < 0.001$  vs. soleus and peroneal). However, the absolute inter-muscle differences were small, FF ranging from 0.6% in the rectus femoris to 2.9% in the sartorius.

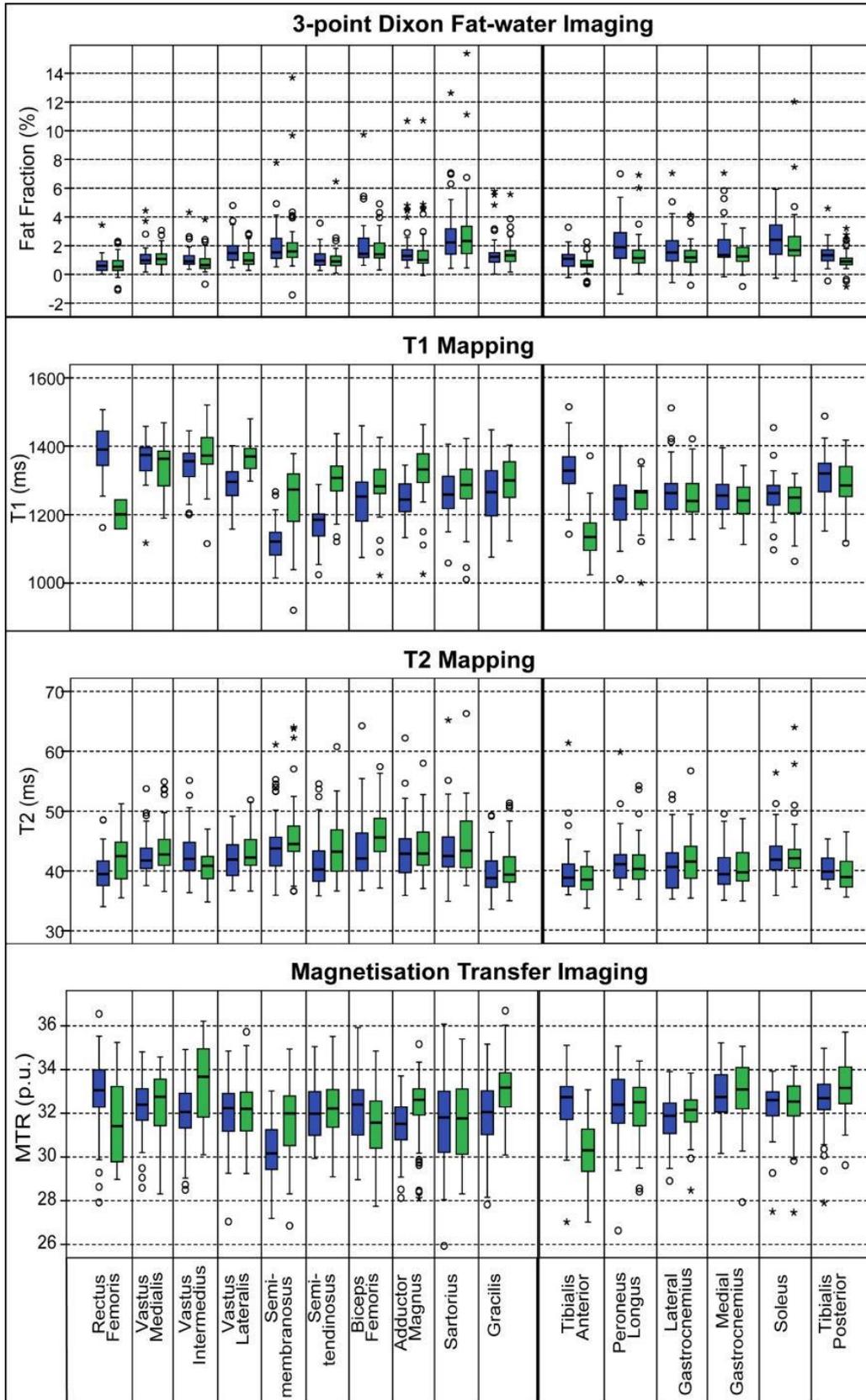
$T_2$  values were also similar in the left and right limbs, confirming independence from  $B_1$  inhomogeneity. Inter-muscle  $T_2$  differences were also significant (ANOVA,  $p < 0.001$  at both calf and thigh-level), with the same muscles (sartorius, semimembranosus and biceps femoris in the thigh; soleus, peroneal in the calf) showing elevated  $T_2$  as elevated FF. Whilst tibialis posterior and tibialis anterior  $T_2$  times were lowest in the calf, consistent with their low FF, gracilis  $T_2$  was lowest despite this muscle's intermediate FF.

MTR showed apparent left-right differences in some regions with lower values for right tibialis anterior, right rectus femoris and left semimembranosus, corresponding to the areas of maximum  $B_1$  deviation. Excepting these ROIs, MTR was similar across all thigh and calf muscles (range 31.7-33.2 p.u.).

Mean  $T_1$  similarly varied between left and right limbs in these muscles, suggesting incomplete  $B_1$  inhomogeneity correction, but was otherwise consistent across the remaining muscles (1240-1370ms).

Fat fraction values obtained from full ROI were consistently higher than those obtained from small ROI (Table 3-7). The overall mean obtained using whole ROI was 3.7% in the thigh and 2.8% in the calf versus for small ROI 1.5% in the thigh and 1.6% in the calf. Combining the large ROI allowed calculation of the total cross-sectional area of muscles, this was  $177 \pm 40 \text{cm}^2$  for muscles at mid-thigh and  $90 \pm 33 \text{cm}^2$  at mid-calf level.

Figure 3-2: Individual muscle ROI values at thigh and calf levels for 47 healthy subjects



Bars indicate median, 25<sup>th</sup>, 50<sup>th</sup> and 75<sup>th</sup> centiles, lines: range, °: minor outlier, \*: major outlier. MTR: magnetisation transfer ratio, p.u. percentage units, blue: left limb, green: right limb

### **3.4.4 Scan-rescan and inter-observer reliability**

Scan-rescan reliability values are shown in Table 3-6, with inter-observer reliability in Table 3-8. Mean values are shown for both summary measures and individual-muscle ROI values, together with scan-rescan and inter-observer ICCs and limits of agreement for both. Intra-class correlation coefficients were 0.84-0.99 for inter-observer and 0.62-0.99 for scan-rescan values, and were generally higher for the summary measures than for the individual muscle values. The limits of agreement were consistently narrower for overall mean values and inter-observer comparisons than for individual ROI values and inter-scan comparisons. Bland-Altman plots for both inter-observer and inter-scan reliability are shown in Figure 3-3.

The limits of agreement for individual muscles are shown in Table 3-11. For scan-rescan measurements there was no evidence of systematic bias, however for inter-observer measurements some muscles had evidence of systematic bias. This was true for area measurements based on whole muscle ROI, suggesting one observer consistently included more voxels at the muscle boundary, which proved true on direct comparison of a sample of ROIs drawn. This resulted in systematically higher fat fraction for the observer with larger ROI, due to partial volume effects with fat tissue adjacent to the muscle border. For those sequences where B1 inhomogeneity was not completely corrected (MTR and T1) there were systematic inter-observer differences for muscles close to areas of remaining inhomogeneity – most notably right vastus medialis. Review of these ROI showed that this was due to systematic variation in ROI placement within the muscle relative to the area of signal inhomogeneity.

In general, inter-observer reliability whether assessed by limits of agreement or intra-class correlation coefficients was better than inter-scan reliability. For fat fraction assessments there was variation in inter-scan reliability between thigh muscles with rectus femoris, gracilis and sartorius showing widest limits of agreement – the smallest muscles. In general reliability was similar for thigh-level and calf-level muscles.

#### **3.4.4.1 Muscle area**

Although there is wide variation in the absolute inter-scan difference in area between muscles, the relative amount is very similar +/- 10% for thigh muscles and +/-5% for calf muscles. There was no systematic bias for inter-scan assessment. The inter-observer reliability of area however shows an overall bias for both thigh and calf with Observer 1 approximately 10% greater than Observer 2 – likely due to including more muscle at the perimeter, but also variation between muscles, suggesting different slice selection. Consistent with this, in the inter-observer comparison of large fat fraction, there was systematic bias with Observer 1 fat fractions approximately 1% higher than

Observer 2, with some variation in this between muscles. This is a consequence of partial volume effect by ROI including muscle boundaries for observer 1. Thus both large ROI fat fraction and muscle area show lower inter-observer ICC than inter-scan ICC.

Table 3-6: Overall reliability table: inter-scan and inter-observer

Parameter	For the overall mean across all ROI				For an individual ROI			
	Mean ± s.d.	IO ICC	IS ICC	IS LOA	Mean ± s.d.	IO ICC	IS ICC	IS LOA
<b>Thigh</b>								
Mean FF small (%)	1.53 ± 0.90	0.93	0.91	-0.50 to +0.27	1.53 ± 1.48	0.79	0.76	-1.25 to 1.01
Mean T1 time (ms)	1279 ± 43	0.95	0.65	-45 to +53	1279 ± 98	0.93	0.79	-103 to 111
Mean T2 time (ms)	42.9 ± 3.1	0.98	0.94	-1.38 to +1.70	42.9 ± 4.8	0.84	0.83	-3.42 to 3.83
Mean MTR (p.u.)	32.0 ± 1.1	0.99	0.87	-1.27 to +1.64	32.0 ± 1.66	0.90	0.71	-2.39 to 2.49
Mean FF large (%)	3.66 ± 1.43	0.52	0.94	-0.83 to +0.49	3.66 ± 2.33	0.59	0.83	-2.28 to 2.13
Total Muscle Area (cm <sup>2</sup> )	227 ± 51	0.84	0.97	-17.2 to +23.1	NA	0.94	0.99	NA
<b>Calf</b>								
Mean FF small (%)	1.58 ± 0.81	0.95	0.89	-0.54 to 0.08	1.58 ± 1.30	0.83	0.62	-2.01 to +1.50
Mean T1 time (ms)	1258 ± 49	0.90	0.64	-64 to +92	1258 ± 78	0.95	0.65	-122 to +142
Mean T2 time (ms)	40.9 ± 3.0	0.96	0.83	-1.70 to +2.34	40.9 ± 3.9	0.86	0.79	-3.11 to +3.80
Mean MTR (p.u.)	32.3 ± 0.9	0.94	0.69	-0.61 to +0.82	32.3 ± 1.5	0.92	0.65	-1.90 to +2.04
Mean FF large (%)	2.79 ± 1.19	0.62	0.84	-1.29 to +1.13	2.79 ± 1.69	0.66	0.80	-2.07 to +1.91
Total Muscle Area (cm <sup>2</sup> )	121 ± 25	0.90	0.99	-5.4 to +5.9	NA	0.98	0.99	NA

IO: inter-observer; IS: inter-scan; ICC: intra-class correlation coefficient; LOA: limits of agreement (+/- 2 s.d. Bland-Altman method); MTR: magnetisation transfer ratio; FF: fat fraction

Table 3-7: Mean quantitative parameters in 15 healthy volunteers and test-retest reliability

Measure	1 <sup>st</sup> Scan Mean ± s.d.	2 <sup>nd</sup> Scan Mean ± s.d.	ICC	Limits of agreement	N
<b>Thigh Level –mean across all ROIs for each subject</b>					
Fat-Fraction (%)	1.36 ± 0.50	1.25 ± 0.58	0.91	-0.51 to +0.28	14
T <sub>1</sub> (ms)	1290 ± 32	1288 ± 30	0.65	-39 to +35	14
T <sub>2</sub> (ms)	42.01 ± 2.28	42.19 ± 2.29	0.94	-1.47 to +1.83	14
MTR (p.u.)	32.23 ± 1.40	32.25 ± 1.19	0.87	-1.63 to +1.67	14
<b>Calf Level– mean across all ROIs for each subject</b>					
Fat-Fraction (%)	1.54 ± 0.65	1.30 ± 0.56	0.89	-0.58 to +0.08*	15
T <sub>1</sub> (ms)	1276 ± 66	1283 ± 56	0.62	-100 to +114	13
T <sub>2</sub> (ms)	39.89 ± 1.75	40.23 ± 2.14	0.83	-1.84 to +2.54	15
MTR (p.u.)	32.80 ± 0.57	32.91 ± 0.44	0.69	-0.67 to +0.89	14
<b>Thigh Level – individual ROI values</b>					
Fat-Fraction (%)	1.32 ± 0.87	1.20 ± 0.82	0.76	-1.25 to +1.01*	254
T <sub>1</sub> (ms)	1282 ± 88	1286 ± 81	0.79	-103 to +111	190
T <sub>2</sub> (ms)	41.90 ± 3.17	42.11 ± 3.27	0.83	-3.43 to +3.83	255
MTR (p.u.)	32.35 ± 1.79	32.40 ± 1.47	0.71	-2.39 to +2.49	224
<b>Calf Level individual ROI values</b>					
Fat-Fraction (%)	1.51 ± 0.93	1.21 ± 0.85	0.62	-1.55 to +0.95*	166
T <sub>1</sub> (ms)	1271 ± 83	1281 ± 79	0.65	-122 to +142	119
T <sub>2</sub> (ms)	39.89 ± 2.60	40.23 ± 2.92	0.79	-3.11 to +3.80*	180
MTR (p.u.)	32.82 ± 1.35	32.89 ± 1.03	0.65	-1.9 to +2.04	160

OM – refers to an overall mean value at thigh or calf level for each patient (15 data points)

ROI – analysing each region of interest individually (300 data points thigh, 180 data points calf)

ICC – intra-class correlation coefficient – test re-test, same observer

LOA – limits of agreement (+/- 2 s.d. by Bland-Altman method)

MTR – magnetisation transfer ratio

Table 3-8: Inter-observer reliability of MRI measurements

Measure	1 <sup>st</sup> Observer Group Mean $\pm$ SD	2 <sup>nd</sup> Observer Group Mean $\pm$ SD	ICC	Limits of agreement	N
<b>Thigh Level – overall mean for each subject</b>					
Fat-Fraction (%)	1.33 $\pm$ 0.50	1.24 $\pm$ 0.41*	0.93	-0.39 to +0.21	15
T <sub>1</sub> (ms)	1293 $\pm$ 33	1287 $\pm$ 34	0.95	-26 to +14	15
T <sub>2</sub> (ms)	41.97 $\pm$ 2.20	42.01 $\pm$ 2.22	0.98	-0.93 to +1.00	15
MTR (p.u.)	32.29 $\pm$ 1.37	32.34 $\pm$ 1.27	0.99	-0.49 to +0.59	15
<b>Calf Level– overall mean for each subject</b>					
Fat-Fraction (%)	1.54 $\pm$ 0.65	1.51 $\pm$ 0.57	0.95	-0.41 to +0.33	15
T <sub>1</sub> (ms)	1275 $\pm$ 63	1276 $\pm$ 66	0.99	-23 to +24	14
T <sub>2</sub> (ms)	39.89 $\pm$ 1.75	39.75 $\pm$ 1.79	0.96	-1.08 to +0.80	15
MTR (p.u.)	32.75 $\pm$ 0.57	32.84 $\pm$ 0.56	0.95	-0.25 to +0.43	15
<b>Thigh Level – individual ROI values</b>					
Fat-Fraction (%)	1.32 $\pm$ 0.91	1.23 $\pm$ 0.91*	0.79	-1.04 to +1.22	281
T <sub>1</sub> (ms)	1289 $\pm$ 88	1284 $\pm$ 94	0.93	-62 to +72	221
T <sub>2</sub> (ms)	41.96 $\pm$ 3.19	42.00 $\pm$ 3.61	0.84	-3.82 to +3.75	280
MTR (p.u.)	32.36 $\pm$ 1.80	32.34 $\pm$ 1.91	0.90	-1.59 to +1.63	248
<b>Calf Level individual ROI values</b>					
Fat-Fraction (%)	1.55 $\pm$ 0.97	1.50 $\pm$ 0.98	0.83	-1.08 to +1.17	172
T <sub>1</sub> (ms)	1269 $\pm$ 81	1269 $\pm$ 81	0.95	-48 to +50	138
T <sub>2</sub> (ms)	39.89 $\pm$ 2.60	39.75 $\pm$ 2.69	0.86	-2.59 to +2.88	180
MTR (p.u.)	32.76 $\pm$ 1.36	32.83 $\pm$ 1.31	0.92	-1.15 to +1.00	175

ICC – Intra-class correlation coefficient

Limits of agreement (+/- 1.96 s.d. by Bland-Altman method)

\* - Evidence of systematic difference between scans/observers ( $p < 0.001$ , Bland-Altman method)

Table 3-9: Inter-scan paired two-tailed t-tests for each modality for each ROI

Scan A vs Scan B	T1	T2	MTR	Dixon
Calf overall	0.686222	0.071556	0.081639	0.117694
Right Tibialis Anterior	0.511788	0.135722	0.023583	0.011669
Right Peroneus Longus	0.923607	0.300702	0.025239	0.285734
Right Lateral Gastroc	0.355663	0.163292	0.232022	0.493376
Right Medial Gastroc	0.495875	0.520808	0.946239	0.09956
Right Soleus	0.12952	0.75991	0.23096	0.757039
Right Tibialis Posterior	0.167532	0.433371	0.014074	0.103283
Left Tibialis Anterior	0.250432	0.002136	0.087645	0.053944
Left Peroneus Longus	0.803196	0.075515	0.151114	0.879453
Left Lateral Gastroc	0.622861	0.322725	0.11571	0.723469
Left Medial Gastroc	0.947682	0.958217	0.393997	0.722407
Left Soleus	0.833965	0.182258	0.122588	0.320411
Left Tibialis Posterior	0.592735	0.021276	0.174642	0.185291
Thigh overall	0.442009	0.222237	0.890963	0.799299
Right Vastus Lateralis	0.413587	0.511788	0.838585	0.27058
Right Vastus Medialis	0.511719	0.923607	0.563317	0.717175
Right Semitendonosis	0.996142	0.355663	0.19361	0.816024
Right Biceps Femoris	0.743214	0.495875	0.889797	0.831138
Right Gracilis	0.098822	0.12952	0.860393	0.289765
Right Adductor Magnus	0.375422	0.167532	0.699227	0.769177
Left Vastus Lateralis	0.849544	0.250432	0.253618	0.754825
Left Vastus Medialis	0.951723	0.803196	0.20126	0.260122
Left Semitendinosus	0.482759	0.622861	0.880393	0.404881
Left Biceps Femoris	0.962006	0.947682	0.232704	0.344563
Left Gracilis	0.756815	0.833965	0.255686	0.141896
Left Adductor Magnus	0.82796	0.045815	0.80465	0.490707

By Bonferroni correction for multiple comparisons for A vs B, for overall thigh/calf (8 measures),  $p < 0.005$  significant

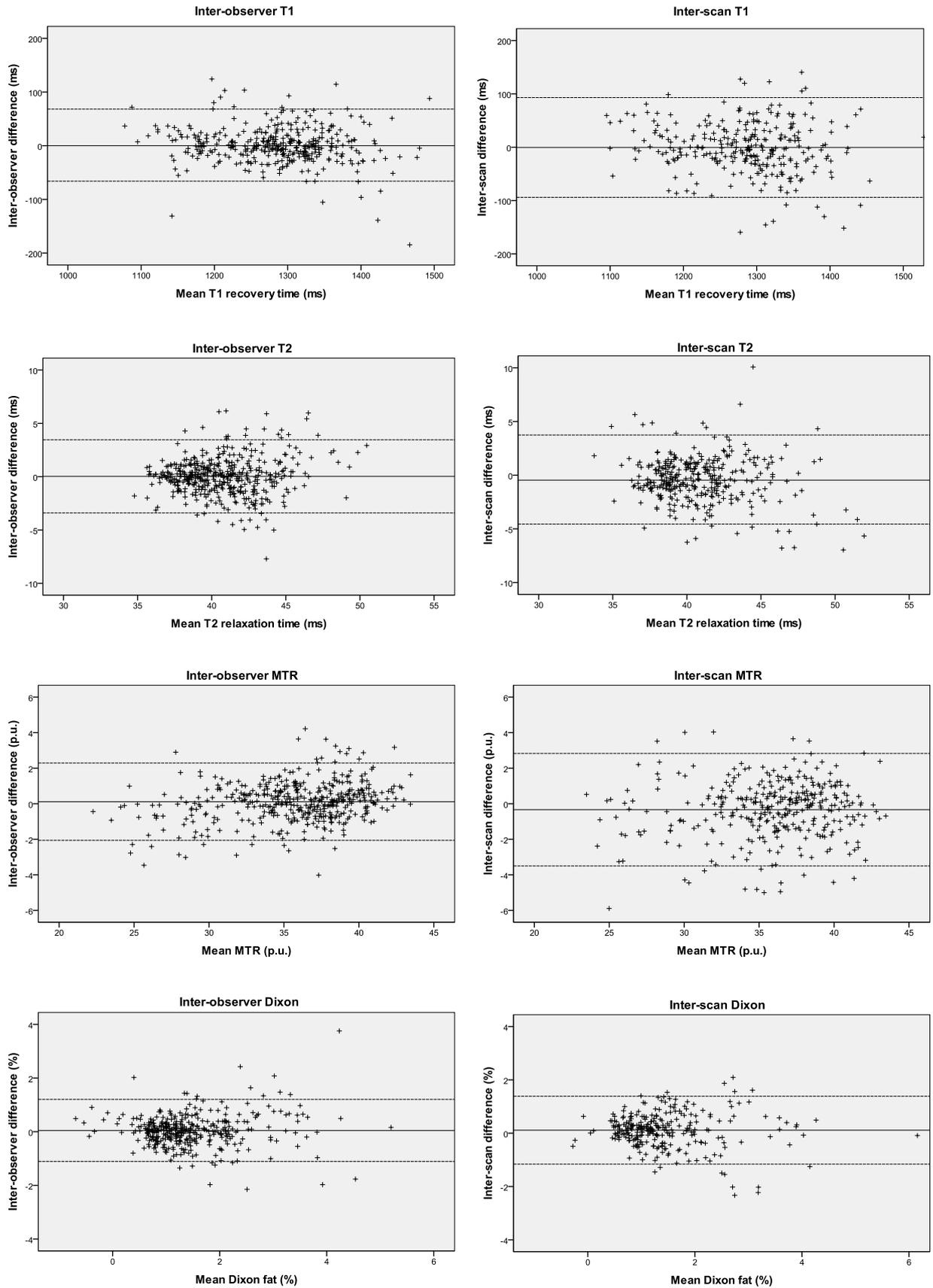
For individual ROIs = 98 measures,  $p < 0.0005$  significant.

Table 3-10: Inter-observer paired two-tailed t-tests for each modality for each ROI

Observer JM vs AF	T1	T2	MTR	Dixon
Calf overall	0.026567	0.425262	0.000389	0.963717
Right Tibialis Anterior	0.943776	0.519198	0.0248667	0.015239
Right Peroneus Longus	0.069205	0.82444	0.2717706	0.072279
Right Lateral Gastroc	0.213473	0.038125	0.0210748	0.308883
Right Medial Gastroc	0.029217	0.09544	2.594E-06	0.020724
Right Soleus	8.2E-06	0.469978	6.154E-06	0.710177
Right Tibialis Posterior	0.233925	0.205691	0.7090662	0.397646
Left Tibialis Anterior	0.737573	0.772459	0.000349	0.056573
Left Peroneus Longus	0.642802	0.615909	0.1594877	0.021507
Left Lateral Gastroc	0.341964	0.291414	2.719E-06	0.174703
Left Medial Gastroc	0.174344	0.167993	0.0002975	0.84026
Left Soleus	2.57E-05	0.030163	0.0030535	0.704276
Left Tibialis Posterior	0.342517	0.320739	0.8132651	0.345205
Thigh overall	1E-05	3.56E-06	7.99E-07	0.005391
Thigh except RVM	0.144547	0.556217	0.556217	0.005364
Right Vastus Lateralis	0.063717	0.086323	1.08666E-07	0.516385
Right Vastus Medialis	3.37E-06	9.87E-10	4.88213E-08	0.147516
Right Semitendonosis	0.022863	0.000573	0.002083682	0.206134
Right Biceps Femoris	0.297453	0.379295	0.029008818	0.628404
Right Gracilis	0.003531	0.046685	0.403574767	0.381144
Right Adductor Magnus	0.053915	0.014977	0.734102468	0.981188
Left Vastus Lateralis	0.321522	0.173646	0.387337283	0.097238
Left Vastus Medialis	0.297033	0.097951	0.995662242	0.013892
Left Semitendinosus	0.001631	0.082517	0.002884189	0.314158
Left Biceps Femoris	0.091009	0.10806	0.002406375	0.457314
Left Gracilis	0.760466	0.807736	0.00011095	0.136857
Left Adductor Magnus	0.023151	0.517997	0.819727905	0.216598

By Bonferroni correction for multiple comparisons for O1 vs O2, for overall thigh/calf (8 measures),  $p < 0.005$  significant. For individual ROIs = 98 measures,  $p < 0.0005$  significant.

Figure 3-3: Bland-Altman plots for inter-scan and inter-observer analyses



Horizontal lines denote the mean difference and 95% confidence limits.

Table 3-11: Limits of agreement (LOA) for each quantitative parameter for each ROI individually

Reliability	IS FF LOA	IS T1 LOA	IS T2 LOA	IS MTR LOA	IO FF LOA	IO T1 LOA	IO T2 LOA	IO MTR LOA
Units	%	ms	ms	p.u.	%	ms	ms	p.u.
L RF	-0.85 to 0.93	-122 to 174	-3.14 to 3.02	-2.02 to 2.76	-0.82 to 0.81	-91 to 99	-2.05 to 1.8	-1.88 to 1.22
R RF	-2.04 to 2.41	X	-2.79 to 2.28	X	-0.76 to 1.22	X	-0.67 to 2.76	-0.52 to -0.52
L VM	-0.66 to 0.33	-184 to 195	-2.87 to 3.77	-2.27 to 2.64	-0.57 to 0.89	-89 to 41	-2.89 to 4.19	-1.79 to 1.18
R VM	-0.88 to 1.18	X	-4.91 to 3.5	0.64 to 0.64	-1.37 to 1.26	-16 to -16	-5.63 to 7.54	-0.66 to -0.54
L VI	-0.81 to 0.56	-83 to 105	-2.44 to 2.61	-2.05 to 2.09	-0.48 to 0.62	-54 to 79	-2.75 to 3.79	-1.63 to 1.09
R VI	-0.84 to 1.06	-19 to 37	-1.91 to 2.67	-3.71 to 2.29	-0.89 to 1.08	-65 to 103	-3.84 to 3.06	-2.12 to 0.35
L VL	-1.04 to 0.7	-93 to 110	-3.01 to 2.95	-1.8 to 2.25	-0.57 to 0.88	-53 to 59	-1.47 to 2.48	-1.23 to 1.2
R VL	-0.57 to 0.29	-43 to 58	-2.18 to 2.62	-2.5 to 1.62	-0.73 to 0.8	-70 to 73	-2.59 to 3.29	-2.06 to 2.23
L SM	-1.72 to 1.23	-78 to 133	-3.13 to 2.84	-1.34 to 2.46	-1.29 to 1.38	-50 to 75	-4.74 to 2.84	-0.68 to 1.78
R SM	-1 to 0.63	-96 to 85	-4.83 to 6.11	-1.43 to 2.01	-1.09 to 1.26	-36 to 72	-4.15 to 3.86	-1.08 to 1.55
L ST	-0.83 to 0.69	-91 to 99	-1.24 to 2.13	-3.62 to 2.81	-0.52 to 0.88	-34 to 47	-2.07 to 1.67	-0.98 to 1.39
R ST	-0.56 to 0.42	-108 to 106	-3.24 to 2.68	-1.89 to 2.35	-0.48 to 0.63	-39 to 36	-2.2 to 2.84	-0.96 to 0.68
L BF	-1.07 to 0.69	-137 to 91	-2.91 to 3.98	-2.99 to 2.62	-0.94 to 0.6	-52 to 73	-3.16 to 1.82	-1.3 to 1.36
R BF	-0.8 to 0.77	-87 to 91	-3 to 4.36	-2.77 to 2.47	-0.72 to 1.06	-45 to 31	-4.51 to 4.57	-1.16 to 0.92
L AM	-0.66 to 0.62	-118 to 111	-2.54 to 4.25	-3.04 to 3.53	-1.06 to 1.12	-63 to 35	-4.81 to 4.11	-2.74 to 2.64
R AM	-1.79 to 1.44	-54 to 51	-3.51 to 5.07	-2.03 to 2.04	-1.78 to 2.71	-51 to 27	-4.84 to 6.86	-1 to 0.38
L Sa	-1.44 to 0.5	-76 to 80	-6.24 to 5.94	-2.21 to 3.23	-1.72 to 2.05	-40 to 104	-6.43 to 4.93	-0.9 to 1.63
R Sa	-1.8 to 1.55	-126 to 144	-5.5 to 5.43	-2.49 to 0.93	-1.13 to 1.62	-78 to 123	-4.77 to 3.82	-1.17 to 2.39
L GR	-1.63 to 0.75	-161 to 112	-3.66 to 3.82	-3.33 to 4.01	-0.56 to 0.83	-39 to 111	-2.37 to 2.07	-1.5 to 3.1
R GR	-2.05 to 1.26	-74 to 117	-3.72 to 3.21	-0.97 to 1.15	-1.74 to 1.04	-63 to 75	-3.81 to 2.46	-0.99 to 1.16
<b>Thigh Total</b>	<b>-1.25 to 1.01</b>	<b>-103 to 111</b>	<b>-3.43 to 3.83</b>	<b>-2.39 to 2.49</b>	<b>-1.04 to 1.22</b>	<b>-62 to 72</b>	<b>-3.82 to 3.75</b>	<b>-1.59 to 1.63</b>
L TA	-1.05 to 0.52	-126 to 146	-1.28 to 4.07	-2.48 to 1.74	-0.33 to 0.49	-7 to 14	-1.68 to 2.36	-0.76 to 0.55
R TA	-1.19 to 0.47	-248 to 210	-3.24 to 1.97	-1.46 to 4.34	-0.24 to 0.31	-135 to 106	-0.64 to 0.82	-1.95 to 1.27
L PL	-1.53 to 1.32	-148 to 186	-2.21 to 3.82	-2.3 to 1.82	-2.18 to 1.22	-54 to 53	-3.89 to 2.13	-0.83 to 0.86
R PL	-3.67 to 1.88	-117 to 177	-6.1 to 5.39	-2.07 to 2.58	-1.93 to 2.48	-36 to 45	-4.3 to 5.07	-1.85 to 1.51
L LG	-1.66 to 1.32	-70 to 161	-3.88 to 4.69	-2.05 to 2.88	-1.1 to 1.06	-67 to 50	-4.57 to 4.19	-1.62 to 1.5
R LG	-0.73 to 0.3	-149 to 128	-3.78 to 4.84	-1.58 to 1.37	-0.39 to 0.83	-47 to 76	-1.94 to 4.34	-0.88 to 0.45
L MG	-1.02 to 0.46	-90 to 69	-2.61 to 3.43	-1.19 to 0.69	-0.81 to 0.68	-39 to 36	-1.81 to 2.23	-0.96 to 1.43
R MG	-1.05 to 0.65	-127 to 154	-3.6 to 4.26	-1.1 to 1.5	-0.94 to 1	-54 to 29	-2.22 to 1.95	-1.22 to 1
L So	-1.31 to 0.4	-76 to 81	-1.94 to 2.59	-0.91 to 0.51	-0.99 to 1.73	-46 to 35	-1.01 to 2	-0.88 to 0.46
R So	-0.82 to 0.65	-131 to 138	-2.24 to 2.56	-1.07 to 1.1	-0.18 to 0.93	-39 to 17	-0.38 to 1.98	-0.67 to 0.28
L TP	-1.22 to 0.66	-124 to 162	-1.78 to 3.38	-1.95 to 1.46	-0.71 to 0.41	-33 to 55	-2.3 to 1.73	-0.61 to 0.84
R TP	-1.19 to 0.64	-179 to 147	-2.19 to 2.2	-1.54 to 2.06	-0.41 to 0.28	-30 to 63	-1.2 to 0.56	-0.44 to 0.75
<b>Calf Total</b>	<b>-1.55 to 0.95</b>	<b>-122 to 142</b>	<b>-3.11 to 3.8</b>	<b>-1.9 to 2.04</b>	<b>-1.08 to 1.17</b>	<b>-48 to 50</b>	<b>-2.59 to 2.88</b>	<b>-1.15 to 1</b>

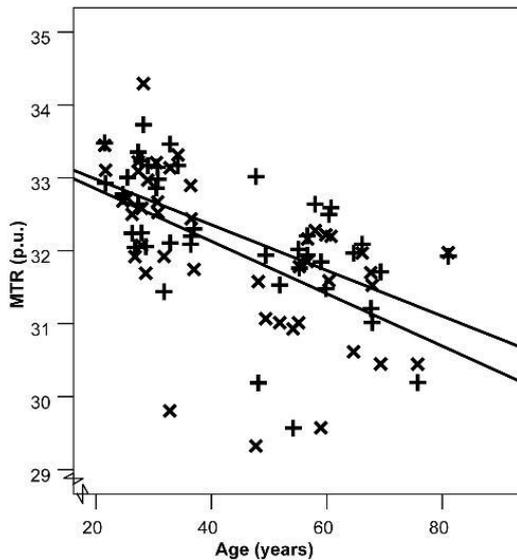
IS: inter-scan, IO: inter-observer, FF: fat fraction, MTR: magnetisation transfer ratio, R: right, L: left, RF: rectus femoris, VM: vastus medialis, VI: vastus intermedius, VL: vastus lateralis, Sa: sartorius, SM: semimembranosus, ST semitendinosus, BF: biceps femoris (long head), AM: adductor magnus, G: gracilis, TA: tibialis anterior, TP: tibialis posterior, PL: peroneus longus, So: soleus, MG: medial head of gastrocnemius, LG: lateral head of gastrocnemius, X: Insufficient data for analysis due to B<sub>1</sub> field inhomogeneity. Thigh and calf totals are the limits of agreement with all muscles at that level combined and correspond to values in Table 3-7.

### 3.4.5 Dependence upon age, gender and weight

Multivariate linear regression modelling the MRI measures at each level against the assumed explanatory variables age, gender and weight results are shown in Table 3-12 for the all-muscle summary measures, and for individual muscles in Table 3-13. There was significant positive dependence of both FF and T<sub>2</sub> upon age at both anatomical levels, and upon weight in the thigh but not calf. MTR showed strong negative dependence upon age ( $p < 0.001$ ) for both thigh and calf (Figure 3-4), illustrating the univariate Pearson correlation between overall muscle mean MTR and age), and significant correlation with weight and notably gender in the thigh. T<sub>1</sub> did not depend significantly upon any demographic parameter, except for an association with weight in the thigh only ( $p < 0.05$ ). Although FF correlated positively with T<sub>2</sub>, and negatively with T<sub>1</sub> and MTR (table 4), the MTR-age correlation remained significant when the other quantitative parameters were included as covariates ( $p < 0.01$  thigh,  $p < 0.001$  calf). As expected, muscle size correlated negatively with age and female gender (Table 3-12, Figure 3-5)

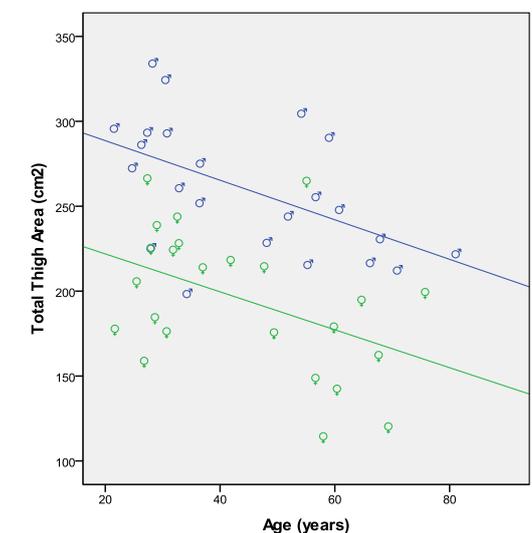
Multivariate linear regression models for individual muscles (Table 3-13) most consistently demonstrated positive correlations between FF or T<sub>2</sub> and weight in the thigh, and negative correlations between MTR and age/gender/weight in the thigh, and age in the calf.

Figure 3-4: Relationship between MTR and age



Overall mean thigh (x) and calf (+) MTR is negatively correlated with subject age ( $p < 0.001$ )

Figure 3-5: Thigh area vs age in males and females



Cross-sectional thigh area reduces with age and is lower in females (green) than males (blue)

Table 3-12: Multivariate regression analysis of muscle MRI measures with subject demographics

	Thigh		Calf	
FF	R=0.58, $p=0.001$		R=0.42, $p<0.05$	
	Co-eff	$p$	Co-eff	$p$
Constant	-0.942	.150	0.573	0.399
Gender	-0.125	.597	-0.329	0.174
Age	<b>0.016</b>	<b>.026</b>	<b>0.014</b>	<b>0.047</b>
Weight	<b>0.025</b>	<b>.003</b>	0.008	0.374
T <sub>1</sub>	R=0.50, $p=0.01$		R=0.43, $p=0.10$	
	Co-eff	$p$	Co-eff	$p$
Constant	1356.9	.000	1278.0	0.000
Gender	15.98	.189	31.34	0.051
Age	-0.519	.149	-0.132	0.770
Weight	<b>-0.858</b>	<b>.035</b>	-0.366	0.468
T <sub>2</sub>	R=0.60, $p<0.001$		R=0.57, $p=0.001$	
	Co-eff	$p$	Co-eff	$p$
Constant	34.44	.000	34.97	0.000
Gender	-0.163	.839	-1.175	0.142
Age	<b>0.074</b>	<b>.003</b>	<b>0.067</b>	<b>0.006</b>
Weight	<b>0.073</b>	<b>.009</b>	0.049	0.064
MTR	R=0.75, $p<0.001$		R=0.61, $p<0.001$	
	Co-eff	$p$	Co-eff	$p$
Constant	35.90	.000	33.707	0.000
Gender	<b>-0.878</b>	<b>.000</b>	0.164	0.485
Age	<b>-0.029</b>	<b>.000</b>	<b>-0.032</b>	<b>0.000</b>
Weight	<b>-0.030</b>	<b>.000</b>	-0.002	0.789
CS Area	R=0.85, $p<0.001$		R=0.79, $p<0.001$	
	Co-eff	$p$	Co-eff	$p$
Constant	208.2	.000	76.6	0.000
Gender	<b>-52.1</b>	<b>.000</b>	<b>-22.7</b>	<b>0.000</b>
Age	<b>-1.408</b>	<b>.000</b>	-0.117	0.431
Weight	<b>1.468</b>	<b>.000</b>	<b>0.834</b>	<b>0.000</b>

$R$  = overall model correlation coefficient; Co-eff = partial regression coefficient;  $p$  = significance level (coefficients with  $p < 0.05$  are indicated in bold type-face).

Table 3-13: Multivariate regression of individual muscle MRI measures with demographic factors

Parameter	FF						T1						T2						MTR						
	Age		Gender		Weight		Age		Gender		Weight		Age		Gender		Weight		Age		Gender		Weight		
Side	R	L	R	L	R	L	R	L	R	L	R	L	R	L	R	L	R	L	R	L	R	L	R	L	
TA	+						-							+				+	+		---				
PL			-						+											-	---				
GL					++			++				--	+	+				++		---	---				-
GM		++		-	+			+	+		-	-	++	++				+		---	---				
SO			-	-					+						-	-				--	-	+			
TP									++				+	+++						-	--				
Calf	+												++						---						
RF		+				+								+		++	-	+		-		---		---	
VM	+				++					+			+	+				+				--		---	
VI					+	+								++		++		+++	--					--	
VL	+				++	+								++	++		++	+	-	--	-	---	--	--	
SM		+		-	++	+	-					--	+			-	+++		---	--			-	-	
ST		+											+			-			-	---	---		--		
BF	+++	+			++	+++	-						+++	+				++	---	---	--	--	-		
AM					++	+	-			+				+			++	++	-	-	-		---		
SA												-							-	-				-	
GR					+	+						-						+	+				--	-	--
Thigh	+				++							-	++					++	---		---		---		

Regression performed on each muscle separately. Calf and Thigh columns refer to mean of all ROIs for each subject, full regression parameters for this are outlined in Table 3-12. The most consistent correlations are: positive between subject weight and thigh muscle F.F., subject age and calf or thigh muscle T2, subject weight and thigh T2, and negative between subject age and MTR in thigh and calf, female gender and MTR in thigh, subject weight and MTR in thigh.

- + = positive correlation,  $0.01 > p > 0.05$
- ++ = positive correlation,  $0.01 > p > 0.001$
- +++ = positive correlation,  $p < 0.001$
- = negative correlation,  $0.01 > p > 0.05$
- = negative correlation,  $0.01 > p > 0.001$
- = negative correlation,  $p < 0.001$

### 3.4.6 Relationship between quantitative parameters

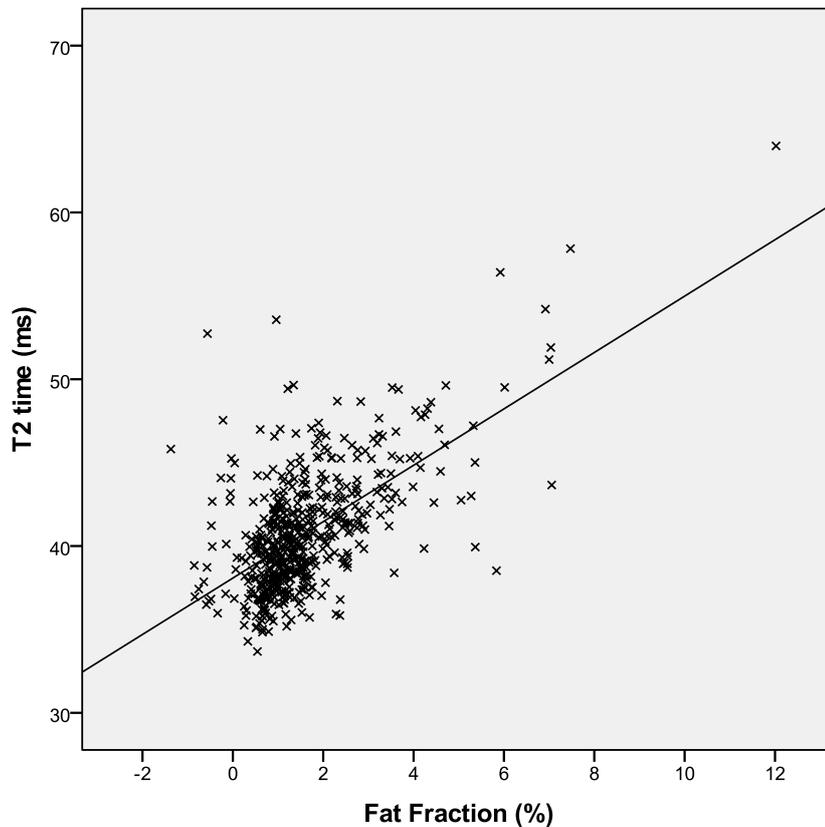
Expected correlations were seen between quantitative parameters within individual small ROI, with fat fraction showing positive correlation with T2 time, negative correlation with T1 time and MTR (Table 3-14). Strongest correlation was seen between T2 time and fat fraction (Figure 3-6). Multivariate regression showed that MTR remained negatively correlated with T2 time independent of correlation with fat fraction or demographic parameters ( $p < 0.001$ ).

Table 3-14: Pearson correlation coefficients between quantitative parameters

<b>Calf</b>	T1	T2	MTR
Dixon FF	-0.28	0.61	-0.30
T1		-0.18	0.42
T2			-0.47
<b>Thigh</b>	T1	T2	MTR
Dixon FF	-0.42	0.62	-0.41
T1		-0.21	0.48
T2			-0.51

Each small ROI from each volunteer is a data point. All correlations significant  $p < 0.00001$

Figure 3-6: Scatter-plot showing correlation between fat fraction and T2 time in the calf



$$R=0.6, p < 1 \times 10^{-53}$$

### 3.4.7 Qualitative imaging

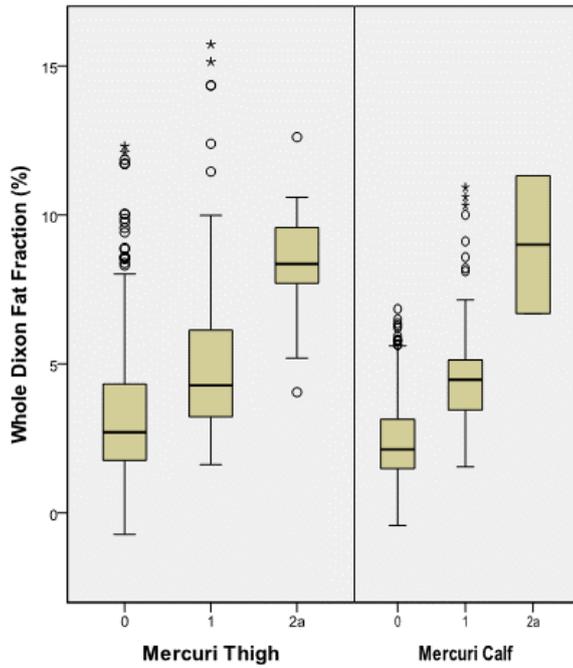
On blinded analysis 18.4% of thigh muscles and 15.0% of calf muscles were scored 1 or 2a on Mercuri grading, most commonly semimembranosus and semitendinosus in the thigh and peroneus longus and soleus in the calf (Table 3-15). Muscles which were graded as “abnormal” on Mercuri grading had increased fat fraction on whole muscle ROI analysis (Figure 3-7) and shortened T1 time, (calf 1204ms vs 1269ms,  $p < 0.001$ , thigh 1193ms vs 1277ms,  $p < 0.001$ ). This confirms that the T1 hyperintensity on qualitative imaging resulting in grade 1 assignment is due to increased fat deposition.

STIR hyperintensity was noted in only 2.8% of thigh muscles (exclusively one of the hamstring muscles or adductor magnus) but in 14.2% of calf muscles, most commonly medial or lateral gastrocnemius (Table 3-15, Figure 3-9). Volunteers with STIR hyperintensity in medial gastrocnemius had a greater BMI but age and gender did not affect the likelihood of STIR hyperintensity (Table 3-16). The small ROI in those medial gastrocnemius muscles with STIR hyperintensity on qualitative assessment had prolonged T2 times and reduced MTR. T1 times and fat fractions were not significantly different (Table 3-16). When combining all calf regions, those with qualitative STIR hyperintensity had significantly prolonged T2 times (Figure 3-10). These findings demonstrate that mild STIR hyperintensity can be seen in calf muscles of healthy volunteers, especially the medial and lateral gastrocnemius, which is reflected in increased T2-times on quantitative imaging.

*Table 3-15: Frequency of Mercuri grades in different muscle groups*

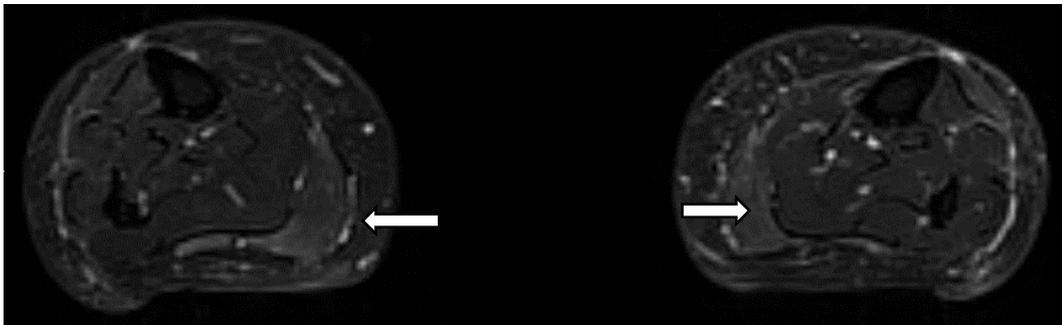
ROI	Grade 0	Grade 1	Grade 2a
Rectus Femoris	100%	0%	0%
Vastus Medialis	79%	21%	0%
Vastus Intermedius	86%	14%	0%
Vastus Lateralis	78%	21%	1%
Semimembranosus	56%	35%	9%
Semitendinosus	62%	36%	2%
Biceps Femoris	75%	21%	4%
Adductor Magnus	83%	17%	0%
Sartorius	96%	4%	0%
Gracilis	89%	11%	0%
<b>Thigh overall</b>	<b>80.3%</b>	<b>18.1%</b>	<b>1.6%</b>
Tibialis Anterior	91%	9%	0%
Peroneus Longus	69%	29%	2%
Lateral Gastroc	99%	1%	0%
Medial Gastroc	79%	21%	0%
Soleus	64%	34%	2%
Tibialis Posterior	100%	0%	0%
<b>Calf overall</b>	<b>83.7%</b>	<b>15.6%</b>	<b>0.7%</b>

Figure 3-7: Box-plot of whole muscle Dixon versus Mercuri grade for thigh and calf



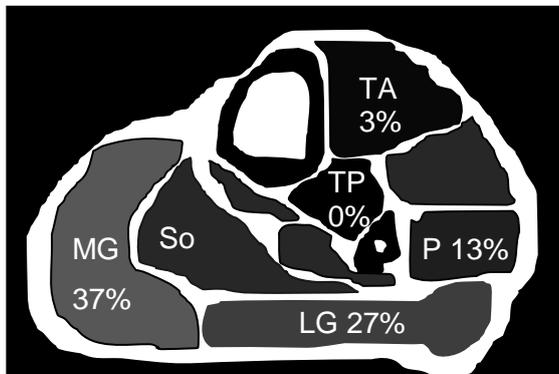
Increasing fraction is measured in calf and thigh muscles from healthy volunteers with increasing Mercuri grades. Bars indicate median, 25<sup>th</sup>, 50<sup>th</sup> and 75<sup>th</sup> centiles, lines: range, °: minor outlier, \*: major outlier

Figure 3-8: STIR hyperintensity in medial gastrocnemius



Mild STIR hyperintensity (arrows) is seen in both medial gastrocnemius muscles in this healthy volunteer

Figure 3-9: Frequency of STIR hyperintensity in calf muscles of volunteers



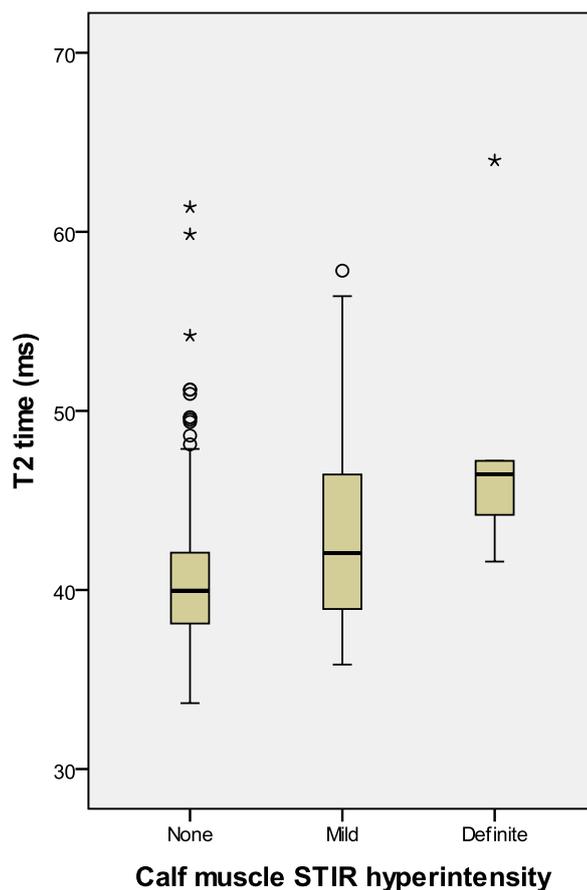
Percentage of volunteers with STIR hyperintensity is shown for each muscle. TA-tibialis anterior; P-peroneal muscle group; LG-lateral head of gastrocnemius; MG-medical head of gastrocnemius; So-soleus

Table 3-16: Demographic and MRI associations of right medial gastrocnemius STIR hyperintensity

Parameter	Normal STIR	Hyperintense STIR	p-value
Female	46%	61%	0.34
<b>Age</b>	<b>38</b>	<b>51</b>	<b>0.01</b>
Height	172	169	0.44
Weight	70	77	0.10
<b>BMI</b>	<b>23.6</b>	<b>26.7</b>	<b>0.03</b>
Dixon FF (%)	1.3	1.7	0.08
T1 time (ms)	1238	1235	0.87
<b>T2 time (ms)</b>	<b>39.4</b>	<b>42.8</b>	<b>0.002</b>
<b>MTR (p.u.)</b>	<b>33.6</b>	<b>32.2</b>	<b>0.002</b>

Quantitative parameters are for right medial gastrocnemius. FF: fat fraction; MTR: magnetisation transfer ratio; STIR: short-tau-inversion-recovery. Significant differences are in bold.

Figure 3-10: Box-plot of T2 times in calf muscles grouped by qualitative STIR classification



Calf muscles with STIR hyperintensity on qualitative imaging have higher T2 time on quantitative imaging. Bars indicate median, 25<sup>th</sup>, 50<sup>th</sup> and 75<sup>th</sup> centiles, lines: range, °: minor outlier, \*: major outlier

### **3.5 Discussion**

These data establish the very good reproducibility of quantitative lower limb muscle MRI obtained using standard acquisition sequences suitable for deployment in NMD treatment trials. The biological sensitivity of the measures is illustrated by demonstrating their dependence upon specific muscle compartment, age, weight and gender in healthy individuals. Even within the narrow range of values seen in healthy volunteers, convergent validity is demonstrated through correlation between the quantitative parameters, and between qualitative and quantitative parameters. Quantitative MRI muscle tissue parameters perform well in terms of the ideal outcome measure characteristics in healthy volunteers: minimal variability and maximal reliability.

#### **3.5.1 Intramuscular fat accumulation**

This study utilised three methods expected to be specifically sensitive to intramuscular fat accumulation: three-point Dixon method derived fat fraction, T1-mapping and qualitative T1-weighted imaging. Whilst T2 mapping and magnetisation transfer imaging are also sensitive to IFA, they are additionally sensitive to changes in muscle water distribution so will be discussed separately.

##### **3.5.1.1 Dixon fat fraction**

Overall a low fat fraction within thigh and calf muscles of healthy volunteers was observed, with an overall mean value for small ROI at thigh and calf level of 1.54% and 1.57% respectively. Inter-subject variation was small with standard deviation 0.90% thigh and 0.81% calf. An overall mean fat fraction of >3.3% in thigh or calf can be considered outside the “normal range” using this protocol. These values are similar to those previously reported in healthy volunteers (see 3.2.1.3.2) and are much lower than values reported in previous patient studies (see 4.2.2.1). Thus measurement of fat fraction meets the first characteristic of an ideal outcome measurement in healthy subjects: a narrow range of values compared with the expected patient values.

##### **3.5.1.1.1 Variation between muscles**

Minor variation between muscles in fat fraction range was demonstrated. The total range of this variation however was small, with the lowest mean fat fraction within the small ROIs of 0.8% in rectus femoris to the highest of 2.8% in sartorius. Generally, the quadriceps had slightly lower fat fractions than hamstring muscles in the thigh and soleus more than tibialis anterior and tibialis posterior in the calf. This is in agreement with a previous study which determined the proportion of fat in the calf of patients with FSHD, which included six control subjects in whom the medial gastrocnemius had the highest fat fraction whilst the tibialis posterior had the least (Kan et al., 2009). In the

present study the highest 95% confidence interval for any muscle was 6%, so values higher than this in any muscle could be considered “abnormal” using this protocol.

### **3.5.1.1.2 Variation between subjects**

Within the narrow range of fat fractions across subjects, this study was able to demonstrate correlation between fat fraction and age in both thigh and calf and between fat fraction and weight in the thigh. Again this variation is very small compared with the range of values already reported in patients. This is consistent with the study of Schwenzer and colleagues who demonstrated higher calf-level FF in an older group of healthy subjects (Schwenzer et al., 2009).

Longitudinal changes which occur in healthy aging is another difficulty with a number of other outcome measures in neuromuscular diseases. Verhamme for example demonstrated progressive reduction of CMAP amplitudes in patients with CMT1A over five years, but also saw reductions in CMAP amplitudes of healthy volunteers and was unable to show that progression was faster in CMT1A than volunteers (Verhamme et al., 2009). Similarly, muscle strength and sensory clinical examination are both known to be age dependent. In this cross-sectional study, regression analysis suggests mean annual change for fat fraction within small muscle ROI of 0.014% or alternatively a 1% increase would occur over approximately 71 years. This means that any significant change over the length of a typical clinical study is unlikely to be due to normal ageing, although this of course needs to be confirmed through longitudinal assessment in phase 2 (chapter 4).

There was no effect of gender on muscle fat fraction within small ROI at thigh or calf level. Interestingly subject weight correlated with thigh but not calf muscle fat fraction, presumably reflecting preferential lipid accumulation in the thigh. These quantitative imaging findings are consistent with <sup>1</sup>H-MRS showing muscle fat increase with weight (Schick et al., 2002; Shen et al., 2008) but not gender (Ortiz-Nieto et al., 2010). The regression analysis showed that a difference in weight of 40kg was associated with a 1% difference in thigh muscle fat fraction. These differences are at least an order of magnitude less than changes in NMDs, and too small to cause significant confounding in longitudinal studies.

There additionally seems to be a differential effect at an individual muscle level of these demographic variables. Fat fraction within biceps femoris appeared most associated with age and weight, whilst sartorius and semitendinosus showed no such association. Fat fraction within quadriceps muscles showed moderate association with subject weight, but inconsistent association with age. This differential susceptibility to

physiological stressors might provide insight into the marked differential susceptibility of muscle to disease processes, the cause of which has been elusive to date.

### 3.5.1.1.3 Small versus whole cross-section ROI

Mean fat fraction was higher with whole cross-section ROI in both thigh (3.70% vs 1.54% small ROI) and calf (2.79% vs 1.57% small ROI) levels. This is likely due to the inclusion of fat which runs along intra-muscular septa, as small ROI were defined to avoid these septa, and also due to inclusion of voxels on the muscle perimeter whose volume includes a proportion of inter-muscular fat. This is reinforced by the observation that whole fat fraction results for the second observer were systematically lower than for the main observer, and that the total cross-sectional area of the ROI was marginally smaller which must logically be at the periphery.

Although not assessed in this study, when whole muscle volumes have been assessed for fat infiltration, slightly larger fat fraction values have been obtained (Hiba et al., 2012), presumably due to an increase in intramuscular fat or an increase in the partial volume effect near muscle origins and insertions. A final interesting observation is that overall the test-retest limits of agreement for full cross-sectional ROI are wider than for the small ROI. This is likely due to variability in decisions to include/exclude voxels at the muscle edge.

Previous authors have variously used small ROI, sub-total ROI, full cross-section ROI and whole muscle volume ROI (see Table 3-1), often without specifying the reason for this decision. The potential advantages and disadvantages are listed in Table 3-17. Ultimately the performance of different types of ROI needs to be tested in reliability and longitudinal studies in patients. It is likely that different ROI may be preferable depending on the disease and research question, but their choice requires justification.

Table 3-17: Advantages of different ROI types

Method	Speed of creating	Atrophy assessment	Reliability heterogeneous pathology	Reliability homogenous pathology	Transferability between maps
Single slice small ROI	++++	-	+	+++	+++
Single slice large ROI	++	+	++	++	+
Single slice whole CSA ROI	+	++	++	+	-
Whole muscle volume ROI	-	+++	++++	+	-

Ratings are subjective opinion of author, scaled best to worst: +++++, +++, ++, +, -

### **3.5.1.2 T1 measurements**

By comparison with three-point Dixon fat fraction, T1 mapping performed much more poorly in healthy volunteers. Despite correction for B1 inhomogeneity, antero-medial muscles in the right thigh had signal dropout artefact, meaning these muscles had to be excluded from analysis in the majority of cases, and those that were not excluded on the basis of visual inspection of the maps, still showed significantly lower values than their right-sided counterparts, which was also seen in right versus left tibialis anterior in the calf. Therefore, in spite of mathematical correction and visual checking, the T1 maps generated by this method have intermuscular variation due to B1 inhomogeneity. Furthermore, the T1 values showed lowest inter-scan reliability of overall mean values judged by intra-class correlation coefficients. The method was not sensitive enough to show inter-muscle variation, nor significant correlations with demographic factors, which were apparent using the Dixon-derived fat fraction. Finally, conceptually, muscle T1 values are more difficult to interpret as although they will be reduced with increased intramuscular fat accumulation, abnormalities in water distribution would be expected to increase T1. This would mean that if both acute and chronic pathologies were occurring simultaneously in a patient, an apparently normal T1 time may result. Compared with T2 and fat fraction measurement, relatively few previous studies have measured T1 time of muscle in patients with neuromuscular diseases. This work in healthy volunteers suggests that for this T1 quantification method at least, few additionally useful data are obtained with T1 mapping.

### **3.5.1.3 Qualitative assessment of T1-weighted images**

The minor variation in fat content of normal muscles, however, is extremely relevant when applying semi-quantitative scales of fatty infiltration on qualitative T1 images such as the Mercuri scale, where the distinction needs to be made between “normal muscle” and “scattered small areas of increased signal” (Mercuri et al., 2002b). As in any area of medicine, there is overlap in the range of “normal” and “abnormal”, as is demonstrated in the 17% of muscles of healthy volunteers which were graded as “abnormal” when the observer was blinded to the presence or absence of disease by including subjects with CMT1A which may have mild abnormalities. The grading of muscles from healthy volunteers as having “scattered small areas of increased signal” was not an error, demonstrated by the reduced T1 time and increased fat fraction observed in the quantitative assessment of these same muscles (Figure 3-7).

The distribution of muscles with most Mercuri grades greater than 0 also matched with quantitative analysis. In the thigh the hamstring muscles were most commonly assigned Mercuri grade 1 and occasionally 2A, and these muscles had higher

quantitative fat fractions than the quadriceps muscles. In the calf tibialis anterior and tibialis posterior had the lowest quantitative fat percentages and also fewest Grade 1/2A Mercuri scores. At both thigh and calf levels there were significant correlations between the qualitative and quantitative scoring. This is consistent with the limited data previously available which demonstrated lower fat fraction in tibialis anterior than soleus. This was attributed to increased intramyocellular lipid concentration (Sakurai et al., 2011). However, the absolute variation in intramyocellular lipid concentration is relatively small (Shen et al., 2008; Tamura et al., 2008), and this would not be detectable by qualitative T1w image assessment. Rather the increased fat fraction is likely due to extramyocellular lipid. On qualitative assessment of the T1w this generally appears as thin streaks of hyperintensity within the muscle, presumably along intramuscular septa, most commonly seen in the soleus and peroneus longus in the calf and hamstring semimembranosus and semitendinosus muscles in the thigh.

These findings have important implications in the diagnostic applications of muscle MRI. The original Mercuri scale utilised in this thesis defines grades as in Figure 3-11 below (Mercuri et al., 2002b)

*Figure 3-11: Original Mercuri grades*

<p>Stage 0: normal appearance.</p> <p>Stage 1: early moth-eaten appearance, with scattered small areas of increased signal.</p> <p>Stage 2a: late moth-eaten appearance, with numerous discrete areas of increased signal with beginning confluence, comprising less than 30% of the volume of the individual muscle.</p> <p>Stage 2b: late moth-eaten appearance, with numerous discrete areas of increased signal with beginning confluence, comprising 30–60% of the volume of the individual muscle.</p> <p>Stage 3: washed-out appearance, fuzzy appearance due to confluent areas of increased signal.</p> <p>Stage 4: end-stage appearance, muscle replaced by increased density connective tissue and fat, with only a rim of fascia and neurovascular structures distinguishable.</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

This paper was concerning patients with Bethlem myopathy, and the described texture of a “moth-eaten appearance” may not be a feature in early stage pathology of all diseases. In a subsequent review paper (Mercuri et al., 2007), Mercuri contracted the

six point scale to four points, but with identical descriptions allowing a direct conversion as follows:

*Table 3-18: Conversion of Mercuri Scales*

Original Mercuri Scale (2002)	Revised Mercuri Scale (2007)
0	1
1	2
2a	2
2b	3
3	4
4	4

Other authors have used slightly different scales, for example Fischer (Fischer et al., 2008) utilised a scale which allowed CT and MRI appearances to be graded using the same five point scale.

Stage 0: normal appearance

Stage 1 (mild): traces of increased signal intensity on the T1-weighted MRI\*

Stage 2 (moderate): increased signal intensity (MRI\*) with beginning confluence in less than 50% of the muscle

Stage 3 (severe): increased signal intensity (MRI\*) in more than 50% of an examined muscle

Stage 4 (end-stage disease): entire muscle replaced by increased signal intensity (MRI\*)

\*Or reduced signal intensity in the case of CT imaging

In both of these scales a similar description is used to distinguish grade 1 from grade 0: “traces” or “scattered small areas” of increased signal intensity. The difficulty in applying this scale is apparent in our publication on the qualitative MRI appearance of lower limb muscles in patients with non-dystrophic myotonia (Morrow et al., 2013), where 8/19 healthy volunteers had at least one muscle graded as grade 1. The volunteer cohort in this thesis had overall 15.8% of muscles Mercuri grade 1 and 0.7% of muscles Mercuri grade 2A. This thesis for the first time is able to correlate these “abnormal” Mercuri grades with quantitatively increased fat, and show a similar distribution.

This raises the question of whether to interpret Mercuri grade 1 as showing scattered small areas of increased signal intensity, greater than that seen in healthy subjects. This increases the subjectivity of the scoring as it relies on the rater having a knowledge of the usual range of signal hyperintensity in each individual muscle, and how this might vary with demographic factors such as age, gender, BMI and level of activity. Another option would be to use quantitative sequences such as three-point

Dixon for diagnostic imaging so that a numerical value can be obtained. The disadvantage of this is that it would be very difficult to ensure that methods were standardised enough to prevent normal values needing to be established on each scanner. The best solution perhaps is to consider both grade 0 and grade 1 as potentially normal, and to include a control group of healthy subjects, especially if the patient group demonstrate relatively mild abnormalities, which is the approach we took in our non-dystrophic myotonia cohort (Morrow et al., 2013).

### **3.5.2 Correlation between quantitative MRI measurements**

It is timely to pause between discussion of measurements sensitive to intramuscular fat accumulation and measurements sensitive to abnormal muscle water distribution to consider the inherent relationship between these measures. T1-weighted images are generally considered insensitive to muscle water changes, but in fact, muscle oedema may result in prolongation of T1 which would result in a slightly hypointense appearance on T1-weighted images, but this relatively small difference is dwarfed by marked hyperintensity of intramuscular fat accumulation. Accurate T1 measurement may be able to quantify this increased T1 with muscle oedema but would be confounded by any chronic intramuscular accumulation, with the associated T1 reduction. Three-point Dixon derived fat fraction would reduce if the volume of water in muscle increased (“diluting” the fat content). This may occur with muscle expansion in acute exercise due to hyperaemia or the swelling associated with rhabdomyolysis but is unlikely to be significant in most neuromuscular diseases. Overall therefore the three methods utilised to quantify chronic muscle pathology are expected to be relatively independent to changes in muscle water.

Conversely the water sensitive measures utilised here are influenced by intramuscular fat accumulation. On T2-weighted images both intramuscular fat and muscle oedema appear hyperintense, motivating the use of the STIR sequence utilising a T1-specific nulling method to suppress high signal from fat. The assumption then is that the remaining hyperintensity is due to muscle oedema, although such findings remain non-specific as to the underlying cause (inflammation, acute denervation, other). Within muscle tissue with fat accumulation, the protons within the water and fat components of each voxel each have different T2 relaxation times. The apparent T2 time based on the dual echo method employed in this study (without any fat suppression or fat water separation techniques included) reflects a combination of the T2 of the water and fat components. Based on this one would predict a strong correlation between measured fat fraction and measured T2 in a muscle – and this was indeed evident with R=0.61 calf, R=0.62 thigh, both  $p < 0.000001$  (see Table 3-14).

Similarly, fat tissue does not display significant magnetisation transfer effects, and the MTR measured within the subcutaneous fat ROI in this study was 0.2p.u. Any intramuscular fat accumulation would therefore be expected to result in a reduction in measured MTR, and indeed in this study (Table 3-14) MTR and fat fraction show a strong negative correlation ( $R=-0.41$  thigh,  $R=-0.30$  calf, both  $p<0.000001$ ).

There are many methods to measure the T2 of the water component of tissue only:

- Using a fat suppression pulse (Willcocks et al., 2014)
- Using  $^1\text{H-MRS}$  at different TEs and modelling the decay of the water peak (Forbes et al., 2014)
- Using a fat-water separation method in parallel to T2 quantification such as the IDEAL-CPMG method developed at UCL in collaboration with GSK (Janiczek et al., 2011)
- Through acquiring data at a large number of echo times (16 or more) to allow estimation of two separate T2 decay constants
- By performing separate fat fraction quantification and T2 mapping, and modelling the T2-fat fraction relationship post-acquisition.

Each of these methods has advantages and disadvantages, a full discussion of which is beyond the scope of this thesis. The key aspect to consider in the interpretation of the quantitative water sensitive measures used in this study are that they are whole tissue T2/MTR measurements and as such sensitive to intramuscular fat accumulation in addition to changes in muscle water distribution. The STIR sequence utilised, although qualitative, is able to assess muscle oedema independent of fat intramuscular fat accumulation.

### **3.5.3 Water sensitive measures**

This study also provides good data to support quantitative assessment of water distribution in muscle as an outcome measure in NMD and important messages in terms of qualitative assessment of this on STIR images. Although both MTR and T2 relaxation times as measured here are influenced by intramuscular fat accumulation in addition to abnormal water distribution, it is envisaged that these measurements will prove both more sensitive to early pathology and more responsive to intervention than fat infiltration. This is supported by previous studies (4.2.2.2) of both T2 relaxation time and MTR and is further supported by observations in this study.

#### **3.5.3.1 T2 measurement**

The overall mean T2 value of  $42.9\pm 3.1\text{ms}$  in the thigh and  $40.9\pm 3.0\text{ms}$  in the calf demonstrates a narrow range compared with expected values in patients and is within

the range of T2 values previously reported for healthy muscle (Table 3-1), though as T2 values are scanner and sequence specific, these should not be considered a universal reference range.

#### **3.5.3.1.1 Variation between muscles**

There were minor variations between T2 time of muscles with the lowest value of 39ms in the tibialis anterior to 45ms in biceps femoris. The mean T2 correlated strongly with the mean fat infiltration within the same ROI (Figure 3-6), and furthermore a similar pattern of differential involvement between muscles was observed (Figure 3-2), which suggests that a good deal of the variation in the T2 time of muscle of healthy volunteers is explained by minor variation in the fat fraction. Conversely the lowest thigh T2 was observed in gracilis, which had intermediate fat fraction suggesting variation between muscles in T2 in addition to that resulting from difference in fat fraction. The observation that left and right mean T2 values are similar for all ROI demonstrates that despite the simplistic technique employed to estimate T2, B1 field inhomogeneity was not a source of artefact.

In the calf the increased soleus T<sub>2</sub> compared to tibialis anterior is consistent with previous results (Gloor et al., 2011; Psatha et al., 2012; Sakurai et al., 2011; Schwenzer et al., 2009) attributed to differing proportions of Type 1 muscle fibres (Johnson et al., 1973) with increased intramyocellular lipid (Sakurai et al., 2011). The differences in total fat fraction between these muscles observed in the present work, suggests that extramyocellular lipid may account for some of this difference. Regardless, for outcome assessment, whilst anatomical specificity far exceeds that provided by non-imaging outcome measures such as myometry (Meldrum et al., 2007) and neurophysiology (Robinson et al., 1991), the inter-muscular differences in T2 measurement are far less than the variation in muscle strength or CMAP observed in healthy controls.

#### **3.5.3.1.2 Variation between subjects**

Muscle T2 showed very similar correlations to demographic parameters as were seen with fat fraction measurements: at thigh level with age and weight, and at calf level with age only, whilst there were no significant gender associations. A similar pattern was also seen when muscles were considered individually. This study therefore failed to identify any age-related changes in T2 independent of the effects of intramuscular fat accumulation, though there may be minor effects if a more sophisticated T2 measurement method, as outlined in 3.5.2 was utilised.

### **3.5.3.2 MTR measurement**

Excepting those muscles for which  $B_1$  deviations were too severe for effective correction, MTRs were consistent with previous calf-muscle studies (McDaniel et al., 1999; Schwenzer et al., 2009). However,  $T_1$  and MTR were in some muscles less consistent comparing left and right limbs, reflecting the increased sensitivity to  $B_1$  variation.

#### **3.5.3.2.1 Variation between muscles**

Despite  $B_1$  correction by an established method (Sinclair et al., 2012b), there were clear left-right differences in MTR values for rectus femoris, semimembranosus and tibialis anterior, which are the areas of greatest inhomogeneity based on our  $B_1$  mapping sequence (mean  $B_1$ : tibialis anterior = 0.88 right, 1.28 left, rectus femoris 0.60 right, 1.24 left, semimembranosus 1.21 right, 0.85 left), indicating that there is a residual artefact due to  $B_1$  inhomogeneity. Apart from these muscles MTR is similar in all muscles of thigh of calf, 31.3-33.0 p.u. by this protocol. Values of MTR of less than 10 p.u. have been reported in patients (Sinclair et al., 2012a), demonstrating that the variability seen within and between healthy controls is minor compared with that expected in disease.

#### **3.5.3.2.2 Variation between subjects**

A novel finding in this study is that MTR of muscle is reduced with increasing age. This observation is robust in that it is true in both calf and thigh muscles (Table 3-12, Figure 3-4,  $p < 0.001$ ) and remains true if fat fraction is included as a covariate, indicating that this reduction of MTR is independent of the slight increase in fat fraction with age noted above. Again the overall decrease with age suggested by this cross-sectional study is modest – from the regression analysis a decrease of 1.0 p.u. over 25-30 years (Table 3-14), however it does indicate that MTR is sensitive to changes in the macromolecular architecture of muscle independent of fatty infiltration. Schwenzer et al. did not demonstrate differences in calf-level MTR in older subjects (Schwenzer et al., 2009). Our contrasting observation may be due to acquisition differences, or the advantage of performing  $B_1$  correction (Sinclair et al., 2012b). MTR was the measure most sensitive to demographic factors, the negative correlation with age being highly significant ( $p < 0.001$ ) for both overall means and many individual muscles, and remaining significant with  $T_2$  and FF included as covariates.

MTR also correlated with increasing weight in the thigh, which is likely attributable to an associated increase in fat fraction. However, gender had a significant effect on thigh muscle MTR: overall 0.90p.u. lower in females by regression analysis. When

examined on an individual muscle basis, the presence of correlation appeared different in right and left thighs, which together with the fact that no similar correlation is apparent at calf level raises the possibility that this effect is perhaps in some way artefactual.

### **3.5.3.3 Qualitative STIR imaging**

Also of note is the observation of apparent STIR hyperintensity on blinded qualitative analysis in a third of all medial gastrocnemius muscles and a quarter of all lateral gastrocnemius muscles (Figure 3-9). Whether medial gastrocnemius is analysed separately or all calf muscles considered together, those with qualitative STIR hyperintensity have significantly prolonged T2-times on quantitative assessment of the same muscles, suggesting by convergent validity that this is a true variation in the normal population rather than artefactual. The average BMI of those with STIR hyperintensity in medial gastrocnemius is significantly higher than those subjects, suggesting that increased muscle strain may be a contributing factor. In any case this mild STIR hyperintensity in the gastrocnemius needs to be recognised as part of the spectrum of findings in normal individuals. The finding of STIR hyperintensity did not correlate with fat fraction, again supporting that these measures of abnormal water distribution are able to show differences in muscle composition independent of fat infiltration.

## **3.5.4 Muscle size measurements**

### **3.5.4.1 Cross-sectional muscle area**

Muscle size measurements, although not the focus of this work, are of interest in order to quantify muscle atrophy in patients. As whole muscle volumes were not imaged, only cross-sectional muscle areas could be calculated in this study. However, other work has shown very strong correlations between cross-sectional area measurements and whole muscle volumes (Miokovic et al., 2012).

#### ***3.5.4.1.1 Variation between muscles***

Cross-sectional area measurements come with the same difficulties present in strength and neurophysiology assessments, i.e. marked between-muscle variation. However, as MRI can quantify the size of individual muscles, for example is able to quantify the size of medial gastrocnemius, lateral gastrocnemius and soleus separately, MRI provides greater anatomical specificity than myometric assessment of ankle plantarflexion, or neurophysiology assessment of the motor function of the tibial nerve. Thus MRI may be able to demonstrate focal atrophy in patients and its size parameters are worthy of assessment in patient studies.

### **3.5.4.1.2 Variation between subjects**

There was significant inter-subject variation in gender, age and weight with predictable influence of these demographic parameters upon muscle area. The mean $\pm$ SD total cross-sectional area of thigh muscles was 227 $\pm$ 51cm<sup>2</sup> and of calf muscles was 121 $\pm$ 25cm<sup>2</sup>. The variation between subjects of the measured size of an individual muscle was even greater with a standard deviation of on average 30% of the mean CSA. Much of this variation could be explained on the basis of subject age, weight and gender. There was significant positive correlation with weight at both thigh and calf level (+1.5cm<sup>2</sup>/kg thigh, +0.8cm<sup>2</sup>/kg calf). Overall females had cross-sectional area 52cm<sup>2</sup> lower in the thigh and 23cm<sup>2</sup> lower in the calf than males independent of the effects of age and weight which were included as covariates. There was also a significant negative correlation with age at thigh level (-1.4cm<sup>2</sup>/year), but interestingly the correlation at calf level was not significant.

These results are consistent with a large body of literature on the effects of aging on muscle strength, and in keeping with a longitudinal CT study which showed a 0.8%/year reduction in thigh muscle cross-sectional area (Delmonico et al., 2009). An MRI study of calf muscle volume did show a 1.7%/year decline in calf muscle size over 10 years. This involved a smaller number of healthy volunteers, and did not control for changes in weight with age, but as a longitudinal study is a direct measurement of the change with age. It therefore seems very likely that measurements of muscle size would be significantly confounded by the effects of ageing in longitudinal studies.

Levels of physical activity were not recorded in the present work so were not included in the regression analysis. However, it is well established that training increases muscle size, and it has been established with MRI that immobility reduced muscle size, so activity level is a significant additional confounder to longitudinal studies using muscle size as an outcome measurement.

### **3.5.5 Reproducibility**

The inter-scan limits of agreement provide a measure of sensitivity to detect meaningful change; e.g. for the thigh level, a change in the overall mean measures in FF, T<sub>2</sub>, T<sub>1</sub> or MTR of +0.28%, +1.8ms, -39ms or -1.63p.u. is a significant change at the 95% level for an individual subject. Rates of change of these with specific NMD progression will be confirmed in future natural history studies, but the detectable change thresholds our data suggest are small compared with cross-sectional disease-dependencies (Gloor et al., 2011; Hiba et al., 2012; Huang et al., 1994; Phoenix et al., 1996; Sinclair et al., 2012a) and are in the range of one-year changes in oculopharyngeal muscle dystrophy (Fischmann et al., 2012).

### **3.5.5.1 Reproducibility of measures of intramuscular fat accumulation**

Fat fraction measurements showed excellent test-retest reliability. The limits of agreement for the different quantitative parameters listed in Table 3-6 and Table 3-7 can be interpreted as the minimum apparent change which could be considered true change rather than measurement error in a longitudinal study, whether natural history or interventional. For example, an increase in the overall mean of thigh fat of greater than 0.29% would be considered significant, whereas for an individual thigh muscle ROI an increase of greater than 1.03% would be considered significant. These values are low enough that a significant change in a measurement such as mean fat fraction might be detectable over a single year in a single patient, something which is not possible with any other current outcome measures in a disease such as CMT1A. Reliability measures for a single muscle are generally lower as would be expected, but still sufficiently accurate to reliably distinguish between degree of involvement in different muscles and detect change longitudinally with slightly higher rates of change.

These figures necessarily apply to the low levels of intramuscular fat seen in the healthy volunteers of this study, but indicate small changes in muscle fat can be detected with the three-point Dixon technique. In an interventional study, the greater number of subjects in each group decreases the difference that would be statistically significant for the group as a whole and these data may be used to aid in sample size calculations.

These results demonstrate that quantitative MRI has excellent reliability and a tight range of values in healthy volunteers so is an excellent candidate outcome measure in NMD. Previous studies have examined the inter-observer or intra-observer reliability by repeating their analysis on the same set of acquired images (see 3.2.4) but this is the first study to assess test-retest reliability of quantitative MRI parameters in muscle across separate scanning sessions. Test-retest reliability is of primary relevance to interventional studies where the outcome measure is generally expressed as a change from baseline to follow-up, and any measurement error will mask any genuine treatment effect.

Inter-scan differences exceeded inter-observer differences as a source of variation, the former potentially driven by small scan-scan position inconsistencies. Compliance with a predefined positioning protocol could improve scan-scan consistency (Fischmann et al., 2014). Mean all-muscle summary measures provide superior reliability to individual muscle measures, an approach which would be appropriate in NMD with diffuse rather than specific muscle involvement.

### **3.5.5.2 Reproducibility of water sensitive measures**

As for fat infiltration quantification, the MRI sequences expected to be sensitive to pathological changes in water distribution — T2 mapping and magnetisation transfer imaging — showed good inter-scan reliability in this study. A two-week interval between test-retest scans was chosen here to encompass any short-term physiological variation in these parameters which would potentially confound detection of long-term changes in a longitudinal study. Previous studies (see 3.2.2.1) have demonstrated that T2-relaxation times, muscle volumes and magnetisation transfer all increase following exercise of muscles to exhaustion, however this effect was already reducing 10 minutes after exercise. Furthermore, other studies failed to show significant difference in either children with dermatomyositis or healthy volunteers 0, 30 or 60 minutes following a standard exercise session (Maillard et al., 2005).

The subjects undergoing the test-retest reliability assessments were not specifically instructed to rest or otherwise prior to scanning, but still had excellent test-retest reliability of T2 measurements, with limits of agreement on repeat assessment for the overall mean in calf or thigh of about 2ms. Together with the published literature this suggests that only repetitive loading of individual muscles has a significant effect on quantitative MRI parameters and that this effect is short-lived. Therefore while it may be sensible to follow the same testing protocol on repeat assessments with regard to timetabling of strength testing and MRI scan, ordinary physical activity should not have a detrimental effect on the reliability of the quantitative MRI assessment.

Factors such as recent exercise (Ababneh et al., 2008; Ploutz-Snyder et al., 1997) or diet (Tamura et al., 2008), known to influence muscle  $T_2$  and fat content respectively, were not explicitly controlled. Nevertheless, the high reproducibility and sensitivity to subtle demographic dependencies demonstrated suggest that in practice such variations may be low, an observation that is important for experimental trials where such factors may be hard to control.

### **3.5.5.3 Inter-observer reliability**

Results from different assessors are prone to systematic bias due to differences in exact strategies for ROI placement; for the small ROI this was due to placement relative to areas of artefact, whilst for the large ROI this was due to decisions regarding inclusion or exclusion of voxels at muscle boundaries. Inter-observer reliability is likely to be worse in patients than in healthy controls due to the heterogeneous distribution of pathology. The practical implications of this are that within a single study a single observer should do all ROI if possible. If this is not practical, there should be precise

instructions as to ROI definition with a phase of training to ensure inter-observer consistency.

In a longitudinal study where change in a measurement over time is the primary outcome, a single observer should do all ROI definitions for the same subject, with direct comparison to the initial ROI placement to ensure the same volume of muscle is assessed at both time points. This ROI placement should occur blinded to time-point to eliminate possible bias. That is, the observer should have a series of acquisitions to define ROI which are all from the same subject, but without any information included which identifies scan order or interval.

#### **3.5.5.4 Methods to optimise reliability**

Whilst inter-scan reliability here was good, it is worth considering potential sources of the variation seen, and considering if measurements could be improved with optimisation of these.

##### ***3.5.5.4.1 Predictable physiological differences during the inter-scan interval***

Examples include recent and long-term diet, training, physical activity immediately prior to scanning, temperature, time of day. Of note, none of these were specifically controlled for in our healthy subjects who underwent repeat scanning, though none would have undertaken significant exercise in the hour prior to scanning. Other studies suggest that acute changes in muscles following exercise return to baseline within 30 minutes, whilst changes in diet may have an effect on intramyocellular lipid content, which makes only a small contribution to total muscle fat fraction. Further studies which quantify these effects or assess whether reliability is improved through controlling these variables would be of interest. Certainly in the context of a clinical trial where strength or aerobic assessments are part of the clinical protocol, it would be preferable for the MRI to occur before these assessments, or after a significant rest period.

##### ***3.5.5.4.2 Random physiological variation***

There is likely to exist true variation in measurements which cannot be predicted or modelled, and therefore cannot be eliminated.

##### ***3.5.5.4.3 Differences due to differences in subject positioning***

This will cause anatomical distortion due to muscle deformation and was qualitatively evident during the reliability phase of the study, most commonly due to the gastrocnemius being displaced by the scanning table to different degrees. Due to the orientation of the lower limbs within the scanner, the majority of distortion will be within

the axial plane whereby it should be possible for ROI on repeat scanning to be defined to contain an equivalent volume of tissue; however it is a source of variability avoidable through a precise subject positioning protocol. This was investigated in collaboration with a visiting radiologist, and the reader is directed to this publication for more information (Fischmann et al., 2014). An alternative way to compensate for this is to acquire and analyse whole muscle volume.

#### ***3.5.5.4.4 Differences due to slice positioning***

Qualitative imaging studies have demonstrated that muscle abnormalities are not uniform along the length of a muscle – not surprising considering the typical clinical findings of distal predominant weakness in neuropathies and proximal predominant weakness in myopathies. It is therefore apparent that if analysis is undertaken of a single slice, as in this thesis, ensuring that this slice is at the same proximal-distal location is critical to reliability. The critical aspect here is in the acquisition, where it is vital that at least one sequence includes clear bony landmarks such as the knee or ankle joint so that slice location on baseline and follow-up scans can be determined with reference to this point. It was fortuitous in this study that one of the sequences which was part of the MTR acquisition included a field of view which encompassed the knee joint, and that inter-slice distance for this sequence was 5mm rather than the 20mm distance for most of the other sequences. Because of variability in the reproducibility of surface anatomy-based positioning as used for the acquisitions in this study (see 6.1.3.2), we selected slices for analysis based on a combination of visual inspection for equivalent anatomy and calculated distance of slices from knee joint (see 2.6.2). We investigated the relative accuracy of surface anatomy versus fixed distance from scout image in collaboration with a visiting radiologist, and the reader is directed to the same publication for more information (Fischmann et al., 2014). In future studies, we employed and recommend improved positioning and slice localisation methods together with reducing inter-slice distance to no more than 10mm to improve these aspects of reliability.

#### ***3.5.5.4.5 Differences due to scanner field strength, manufacturer, model, coil, software, exact sequence parameters***

All of these parameters can and do affect quantitative measurements to some degree. In this healthy control reliability study performed over a two-week interval, there were no such changes, but in follow-up studies of longer duration, this can introduce significant systematic bias, which was indeed noted in T2 measurements in this project following a software upgrade towards the end of the baseline patient scanning phase (see 2.4.4.3.1). Multicentre trials are likely to take place using scanners from multiple

manufacturers, and it is important that rigid processes to ensure cross-site standardisation are employed. We have successfully applied this using fat fraction assessment across four sites in LGMD2I (Willis et al., 2014). Other researchers have recently published quantitative MRI protocols for boys with DMD standardised across multiple sites (Forbes et al., 2014). Ultimately however, for an outcome measure it is the longitudinal change measurement which is most critical. Taking the example of a weight loss intervention study, a systematic difference in scales between sites of 2kg is of less importance if each subject uses the same scales throughout the study. Similar MRI trials may be designed so that all assessments for an individual participant occur at the same study site, but consideration must be given that one of the factors above may change during the course of the study beyond the control of investigators. It is for this reason that for natural history studies the inclusion of healthy controls is critical, whilst in clinical trials the placebo arm together with adequate randomisation and blinding will mitigate against any systematic bias which may otherwise occur following for example a scanner software upgrade. It is worth noting also that some methods such as the three-point Dixon method are likely to be more robust in the face of such alterations, and a more sophisticated T2 mapping method, using 16 echoes for example, may be more robust than the dual-echo method employed in this study.

#### ***3.5.5.4.6 Differences due to image post-processing/map generation***

This should not have been an issue in this study as all post-processing occurred using the same algorithms at the same time (Dr Sinclair, acknowledgements). However, in multi-centre studies centralisation of these aspects to a single site would eliminate any variability related to this. If this is to occur locally, procedures will be required to ensure this step does not introduce systematic bias between study sites, as always most crucially the process should be identical for multiple scans of the same participant.

#### ***3.5.5.4.7 Differences due to selection of slice for analysis***

As discussed above (3.5.5.4.4), this is crucial due to proximal-distal gradients of pathology within muscle. Ideally the scanning will be acquired with the image blocks centred in the same position, but if this did not occur, as long as the slices obtain overlap, a slice common to test and retest scans can be chosen for analysis post-acquisition. How closely these will correspond will depend on the inter-slice distance (half this distance will be the maximum proximal-distal distance between baseline and follow-up slices), and how accurately slice location can be determined. In future studies we have utilised at most a 10mm inter-slice distance and where possible a 5mm slice distance, and further analysis of this latter “3D Dixon” data set will be able to analyse the effect of these different methods in detail.

#### **3.5.5.4.8 Region of interest definition within the slice**

As this work shows, there is variability introduced by different observers drawing regions of interest on identical images. Other studies have examined intra-observer reliability and found this to be superior to inter-observer reliability. To maximise reliability, one observer should be used wherever possible, certainly for longitudinal assessments within a single subject. If whole muscle ROI are utilised, clear rules with regard to defining ROI boundaries are important, for example are only voxels appearing to wholly lie within the fascia included, or should voxels with at least 50% volume within the fascia be included. The choice will depend on the sequence and research question, but consistency is important. For some muscles such as the hamstrings, the inter-muscular boundary is indistinct – and in case of the portions of the vasti, arbitrary. If individual muscle ROI are drawn, it is important to be able to draw ROI with direct comparison to the other ROI in the same subject to ensure consistent decisions are made with regard to anatomy.

#### **3.5.5.4.9 Extraction of data from ROI and subsequent data processing**

Once maps are created and ROI are defined, the data extraction is an automated process. There does however need to be a process where maps are checked for artefact within the ROI so that these values can be excluded. The data may then require transfer to a platform which allows analysis. As far as possible these processed should be automated – and certainly not involve manual data transcription which may introduce transcription errors. Finally, there should be a process to cross-check the final data set back to the original data in the case of outliers and also of a random sample to make sure there are no analysis/processing errors which could have a major effect on reliability.

#### **3.5.5.4.10 Optimising reliability summary**

Optimising reliability necessarily involves consideration of all the above potential sources of measurement variability in a systematic way and eliminating or reducing each of them as far as possible. Additional variation in longitudinal change introduced by poor test-retest reliability will increase the standard deviation in the change measurement, resulting in a significant reduction in outcome measure responsiveness, critical to study power.

### **3.5.6 Feasibility/Study limitations**

To facilitate widespread applicability, sequences were chosen to be readily implemented on standard scanner hardware and software, and with which comprehensive anatomical coverage in a practical examination time was possible. This necessarily limited the measurement sophistication, e.g. multi-echo T<sub>2</sub>

measurement sequences allowing analysis of multiple  $T_2$  decay components (Ploutz-Snyder et al., 1997) did not meet the requirements of ready availability and anatomical coverage vs. acquisition times. However, we showed that these sequences chosen are adequate to provide sensitive and reproducible measures of FF,  $T_2$  and MTR relevant to muscle pathology.

A limitation of quantitative MRI in the legs is the inherent  $B_1$  inhomogeneity, particularly at field strengths of 3T and higher. While the dual-contrast TSE  $T_2$ -relaxometry sequence and Dixon FF measurements were reasonably robust to this, MTR and  $T_1$ -relaxometry data were more affected despite our adoption of  $B_1$ -correction strategies. Nevertheless, MTR proved highly sensitive to age, weight and gender dependencies, although the  $B_1$  correction proved insufficient for meaningful  $T_1$  and MTR quantification in regions of maximum  $B_1$  deviation: worse anteriorly in the right thigh using our scanner-coil configuration. With the methods used,  $T_1$  was the least reproducible measure, the least sensitive to demographic variations, and did not add explanatory power for these factors. While this low sensitivity may reflect  $B_1$ -related measurement error rather than fundamental insensitivity, we conclude that muscle  $T_1$  may not be useful as a NMD outcome measure.

It should be noted that  $T_1$ ,  $T_2$ , FF and MTR values may be partially method implementation dependent and the values presented here cannot be considered an absolute reference range if alternative methods, scanner platforms or field strengths are used, and multi-centre patient trials should confirm inter-centre value comparability during study initiation.

### **3.6 Conclusions**

These data demonstrate significant strengths of quantitative MRI as an outcome measure in neuromuscular diseases. We have shown tight range of normal values, minimal variation with subject demographics and excellent test-retest reliability. The data also reveal normal variation between muscles of fat fraction and  $T_2$  relaxation times which correlate with qualitative  $T_1$  and STIR assessments and should be taken into consideration when making these assessments.

There are of course a number of different quantitative methods (see Table 1-1) to measure fat infiltration in muscles which have already been applied to neuromuscular diseases, including  $T_1$  measurement, measurements based on  $T_2$  mapping, proton spectroscopy, a dual-echo dual-flip-angle spoiled gradient-recalled MR imaging technique, as well as those based on the Dixon technique such as this thesis. All of these techniques have shown good correlation to clinical markers of severity, muscle

strength and muscle biopsy, which demonstrates validity as an outcome measure (see 4.2.4).

Similarly, there are variations in methods to measure T2 relaxation time, with the method chosen here using two echo times theoretically less precise than those utilising a greater number of echo times. However, value as an outcome measure should be judged upon validity, reliability and responsiveness (Kirshner and Guyatt, 1985) balanced against practicality in terms of scan duration, hardware and software requirements and time taken for analysis. The protocol presented here is intended to be able to quantify both early changes of abnormal water distribution which would be potentially responsive to intervention and also the progressive changes of fatty infiltration to demonstrate reduced progression of disease in a clinical trial, while also being practical to implement and analyse on standard MRI hardware in a reasonable length of time. This quantitative neuromuscular MRI protocol was considered appropriate to progress to use in cross-sectional (chapter 4) and longitudinal (chapter 5) studies in patients.

## **4 Cross-sectional quantitative MRI in neuromuscular disease**

### **4.1 Introduction**

Cross-sectional patient studies are the crucial second step of outcome measure development. The measurement needs to be shown to be sensitive to “abnormality”, that is significantly different in patients than in controls. Distinguishing subjects with disease from healthy subjects is key for measurements undertaken for diagnostic purposes, and determines the sensitivity of a diagnostic test. Specificity, which is distinguishing patients with one disease from those with another, is also very important in diagnostic tests, however is a less important quality for outcome measures. Indeed, measurements being consistently abnormal across multiple diseases may be seen as an advantage since this allows these measurements to be applied as outcome measures across many diseases, in comparison with a disease-specific measure which is necessarily limited in its application.

#### **4.1.1 Validity**

An abnormal measurement however is insufficient for an outcome measure; it must also be shown to have validity. Direct measures of mortality or morbidity are preferred by regulatory bodies such as the FDA but “surrogate” outcome measures are allowed in some circumstances. Surrogate outcome measures are most often used as an intermediary endpoint, for example anti-hypertensive medications may now be licensed on a blood pressure lowering endpoint, which has been extensively shown to have long-term benefits in reduction of mortality from cerebrovascular and cardiovascular disease. To be accepted as a surrogate, the biomarker must be shown to have a critical role in pathogenesis. For example, deoxysphingolipid levels are elevated in Hereditary Sensory Neuropathy type 1 (HSN1) caused by mutations in serine palmitoyl carnitine transferase. Treatment with serine has been shown to reduce plasma concentrations of deoxysphingolipids (Garofalo et al., 2011), but it remains to be shown that this translates to patient benefit.

Muscle MRI cannot be considered a “true” outcome measure as it is not a direct measure of patient mortality or morbidity. However, neither is it a surrogate outcome measure in the usual sense, as it is not intended in general as a predictor of future morbidity or mortality, but rather as a more reliable and sensitive measure of current morbidity than direct measures of function. A similar example might be measurement of spirometry parameters such as FEV1 in chronic obstructive pulmonary disease (Ben Saad et al., 2008). This is a highly relevant functional measure with established correlations to exercise tolerance and hospital admissions, both of which are direct morbidity measures. However, if FEV1 can be more reliably measured than a six-

minute walk test due to greater test-retest reliability it would prove a more sensitive outcome measure. Usually such “proxy” outcome measures would be utilised as primary outcome measures in phase 2 clinical trials, or secondary outcome measures in phase 3 clinical trials. However, they might be accepted by regulatory bodies as a primary outcome measure in a phase 3 trial through an orphan drug programme.

Validity is simply defined as the degree to which an instrument measures what we intend it to measure (Roach, 2006). In the present context this means the degree to which our MRI measurement is an accurate measurement of disease severity, and longitudinally of disease progression. If we wish the outcome measure to be considered as evidence for drug licensing, validity should be with reference to patient function (or survival), but for phase 2 clinical trials, validity with reference to disease severity may be adequate. Validity is not an inherent characteristic of the measurement instrument itself and can be determined only in relation to a particular question in a defined population (Roach, 2006). For example, fat accumulation may be a valid measure of disease severity in skeletal muscle or the liver, but not in cardiac muscle or the brain. Even for skeletal muscle IFA may be a valid measure of disease severity in some neuromuscular diseases but not others. Finally, for a particular neuromuscular disease such as Duchenne muscular dystrophy, MRI measured IFA may be a valid measure for a long term intervention such as gene therapy, but not valid for short term assessment of the same intervention.

There are many dimensions in which validity can be considered and demonstrated (Kirshner and Guyatt, 1985; Roach, 2006). Face validity is the extent to which a test appears to measure what it is intended to measure. As measures of IFA, three-point Dixon fat quantification would have superior face validity than T1-mapping, since it is a direct attempt to measure the underlying disease process. Content validity is the degree to which a test includes all the items necessary to represent the concept being measured. This differs by population: for example, it may be that in some diseases muscle atrophy is the primary cause of weakness whilst in another it is intra-muscular fat accumulation. However, if both processes are involved a measure which incorporates both would have greater content validity. Similarly, a measure like T2 which is sensitive to both acute and chronic pathology may have higher content validity for overall disease severity. Face and content validity cannot be directly examined experimentally, and are therefore considered by some as lower levels of validity (Roach, 2006).

Criterion and construct validity can be examined experimentally. For criterion validity, outcome measure validity is tested by comparing the results of the outcome measure to a gold standard or criterion test (Kirshner and Guyatt, 1985). In the case of

neuromuscular disease, patient function is the gold standard for treatment efficacy. Correlation to strength measurements would be another example of a means to assess criterion validity of quantitative MRI. Construct validity reflects the ability of a test to measure the underlying concept of interest to the researcher (Roach, 2006). This can be assessed with the known groups method, by assessing if the measurement differs in groups known to differ in construct of interest. The simplest assessment here is to compare measurements in a patient group to a healthy control group, although comparing results between different patient groups would also provide construct validity. Construct validity can also be assessed by looking for convergent validity. This is demonstrated when results of the test being examined are highly correlated to results of a test thought to measure similar or related concepts. For neuromuscular MRI this might be assessed by comparing the measurements from novel quantitative MRI methods to those obtained with the established qualitative MRI methods. For example, quantitative fat quantification using three-point Dixon could be compared with Mercuri grading of qualitative T1-weighted sequences.

A cross-sectional study including more than one patient group and matched controls, including novel and established MRI measures, and including additional clinical measures of disease severity and function, provides ideal data to assess outcome measure validity.

## **4.2 *Background literature***

### **4.2.1 Qualitative MRI studies in CMT1A/IBM**

With the exception of the pilot study from our Centre of the magnetisation transfer ratio of calf muscle in patients with CMT1A (Sinclair et al., 2012a), there are no previous quantitative muscle MRI publications in CMT or IBM. Publications reporting the appearances of lower limb muscles using qualitative MRI sequences are relevant in predicting the distribution and type (acute vs chronic) of pathology present, and therefore ensuring that the quantitative sequences applied are likely to adequately cover the pathology present. In this way qualitative MRI may be considered the precursor to a quantitative study, and we have used qualitative MRI in studies performed concurrently to this thesis in diseases not previously studied with muscle MRI: non-dystrophic myotonia (Morrow et al., 2013) and congenital myasthenic syndromes (Finlayson et al., 2016).

#### **4.2.1.1 Published reports in CMT1A**

There are six published papers regarding the lower-limb muscle appearance on qualitative MRI in CMT1A (Berciano et al., 2008, 2010; Chung et al., 2008; Gallardo et

al., 2006; Kim et al., 2015; Pelayo-Negro et al., 2014) reporting a total of 69 patients. These studies have all included T1w images, and report with this sequence a distal predominant accumulation of intramuscular fat in the lower limbs, consistent with the clinical phenotype. In the mildest patients, calf musculature is normal with only foot musculature affected (Berciano et al., 2008; Gallardo et al., 2006). In the most severe patients, calf musculature shows end stage changes with near complete fat replacement of calf muscles. In subjects with an intermediate clinical severity and calf involvement on MRI, there is a predilection to the anterior and lateral compartments with a distal gradient of involvement within these muscles (Chung et al., 2008; Gallardo et al., 2006). Only one study included sequences sensitive to acute denervation (T2-weighted and GAD), which showed acute changes in a subset of the subjects (Gallardo et al., 2006).

Taken together these studies would predict that the calf musculature would be the most informative scanning area in CMT1A patients overall, although in severely affected patients the thigh musculature and in mildly affected patients the foot musculature, may prove more helpful. Sequences sensitive to fat infiltration are confirmed by all studies to be likely to be informative, whilst one study shows qualitative sequences sensitive to abnormal water distribution are abnormal in a subset of patients, which may also be worthy of quantitative assessment.

#### **4.2.1.2 Published qualitative MRI reports in IBM**

Five studies have reported qualitative MRI in skeletal muscle of IBM patients (Cantwell et al., 2005; Cox et al., 2011a; Phillips et al., 2001; Reimers et al., 1994; Sekul et al., 1997), the largest of which enrolled 32 patients (Cox et al., 2011a). A selective pattern of muscle involvement has been reported, where similar to the clinical distribution of weakness, there is preferential intramuscular fat accumulation of flexor digitorum profundus within the forearm (Cantwell et al., 2005; Cox et al., 2011a; Phillips et al., 2001; Sekul et al., 1997); with quadriceps predominantly affected in the thigh (Cox et al., 2011a; Phillips et al., 2001; Reimers et al., 1994). Some authors report relative preservation of rectus femoris within the quadriceps (Cox et al., 2011a; Phillips et al., 2001). Within the lower leg, the medial head of the gastrocnemius has maximal intramuscular fat accumulation (Cox et al., 2011a; Phillips et al., 2001; Reimers et al., 1994); a feature hidden clinically because the soleus and the lateral gastrocnemius with lesser involvement compensate in ankle plantarflexion strength. STIR sequences have been performed in three studies with hyperintensity reflecting active inflammation commonly seen, but in a smaller number of muscles than are affected by fat accumulation (Cantwell et al., 2005; Cox et al., 2011a; Sekul et al., 1997).

These publications suggest that both fat and water quantitative MRI sequences will be of interest in IBM patients, with both thigh and calf musculature in the lower limbs likely to be involved, though the thigh involvement may be functionally relevant in a greater proportion of patients.

#### **4.2.2 Studies of quantitative MRI in neuromuscular diseases**

Although there is a paucity of previous literature on quantitative MRI in CMT1A and IBM, there is an increasing literature in a range of other neuromuscular disorders. The first assessment should be whether the measure is abnormal in patients versus the healthy population.

##### **4.2.2.1 Quantitative MRI measures of chronic muscle pathology**

A number of studies have used MRI to quantify chronic muscle damage in neuromuscular disease. Pathologically chronic muscle damage whether due to primary muscle disease or denervation results in atrophy, fat infiltration and fibrosis of muscle to varying degrees. Whilst any MRI sequence can measure muscle size, those with highest resolution and best tissue contrast, especially for fascial boundaries, will do so most accurately. There are a number of fat quantification methods, as discussed in chapter 3, whilst a robust way to quantify muscle fibrosis with MRI remains elusive.

Some investigators have used a semi-quantitative method to estimate fat infiltration where a standard qualitative T1w sequence is segmented based on signal intensity – as fat as higher signal intensity than healthy muscle. This allows calculation of residual muscle volume, fat-infiltrated muscle volume, total muscle volume and hence the percentage of muscle which is fat infiltrated. This has been performed in myotonic dystrophy (Hiba et al., 2012), DMD (Torriani et al., 2012) and spinal muscular atrophy (Sproule et al., 2011a) and demonstrated abnormalities compared with controls (Hiba et al., 2012; Torriani et al., 2012) or between patient subtypes (Sproule et al., 2011a). The potential advantage of this method is the easy availability of a T1w sequence on any scanner, but as the resultant signal intensities are relative, standardisation across different platforms is difficult. B0 field inhomogeneity, especially at higher field strengths (3T), confounds analysis; the method relies on presence of normal muscle as a reference which may not be available in more severely affected patients, and the binary result for each voxel (normal/fat) may reduce sensitivity if fat infiltration is diffuse. Some of these issues may be addressed by estimating the underlying T1 recovery time – i.e. T1 mapping. This was applied to DMD in 1994 (Huang et al., 1994), but subsequent studies have preferred other methods of fat quantification.

T2 relaxation time is significantly greater for fat than healthy muscle, meaning muscle T2 measurement is also a potential surrogate measurement for fat infiltration. Whole

tissue T2 (i.e. without any attempt at separation or suppression of the fat signal) has been shown to be elevated initially in a mixed population of muscular dystrophies (Phoenix et al., 1996), and subsequently in oculopharyngeal muscular dystrophy (Gloor et al., 2011), DMD (Huang et al., 1994; Kim et al., 2010a; Triplett et al., 2014), and Bethlem myopathy (Triplett et al., 2014). One of the disadvantages of T2 measurement is that T2 is also influenced by changes in muscle water (e.g. inflammation or acute denervation), and indeed has been used to measure this effect in juvenile dermatomyositis (Maillard et al., 2004b). By utilising a larger number of echo times in T2 modelling, a bi-exponential fit can be calculated, which allows an estimation of contribution to T2 relaxation of fat and water components, and thence a fat fraction to be calculated. This has been applied to patients with FSHD (Kan et al., 2009).

The three-point Dixon fat-water separation method is the most commonly used technique of fat quantification, and has been used to show increased fat content versus healthy controls in myotonic dystrophy (Hiba et al., 2012), oculopharyngeal muscular dystrophy (Gloor et al., 2011), Duchenne muscular dystrophy (Triplett et al., 2014; Wren et al., 2008), Bethlem myopathy (Triplett et al., 2014) and in LGMD2I (Willis et al., 2014). Many of these studies have been multi-centre, proving the method robust to use across different sites and platforms.

Another common but not universal feature of chronic muscle disease is reduction in muscle size – atrophy. This can be assessed with almost any MRI sequence but there are some difficulties in its application. Two main approaches are possible: measurement of a whole muscle (or limb) volume, or measurement of muscle cross-sectional area. The former is labour intensive unless automatic segmentation protocols are developed, and the definition of muscle boundaries at their proximal and distal ends may be challenging, whilst the latter is dependent upon exact positioning in the superior-inferior direction and can be affected by tissue deformation (Fischmann et al., 2014). Within neuropathies, volumetric assessment has been applied to foot muscles in diabetic neuropathy (Andersen et al., 2004) and differences in total muscle volume have been shown in SMA (Sproule et al., 2011a). In muscle diseases differences in total muscle volume or CSA are less marked with no significant difference seen in CSA of calf muscles in mixed NMD (Phoenix et al., 1996), DMD (Torriani et al., 2012; Wokke et al., 2014) and myotonic dystrophy (Hiba et al., 2012), although residual muscle volume measured using some method of fat quantification above is reduced. The other difficulty in using size-based measurements as outcome measures is the confounding effects of age, gender and physical activity/inactivity – which are common of course to direct strength measurements.

Overall these studies demonstrate that MRI measures sensitive to intramuscular fat accumulation are abnormal in patients with a broad spectrum of neuromuscular diseases. Results from MRI measurement of muscle size are more varied, but are also significantly different to controls, especially in neuropathies.

#### **4.2.2.2 Quantitative measures of acute muscle pathology in neuromuscular diseases**

There are overall fewer studies which have attempted to quantify acute pathology in neuromuscular diseases. Measures of abnormal muscle water distribution are confounded by the effect of fat: T2 is increased in acute denervation and acute muscle inflammation but even more so with fat infiltration; similarly MTR is reduced in acute pathology, but near zero in fat tissue. The situation is analogous to qualitative imaging where a fat suppressed T2 weighted sequence such as STIR is used to separate acute from chronic muscle pathology. A semi-quantitative STIR method using a saline phantom as reference has been used to evaluate abnormal muscle water distribution in hypokalaemic periodic paralysis (Jurkat-Rott et al., 2009), whilst studies of focal neuropathies have used nearby healthy muscle as an internal reference.

Other authors have utilised quantitative methods. In myotonic dystrophy type 1, the T2 of “residual muscle tissue” segmented based on T1w voxel signal intensity, showed increasing tibialis anterior muscle T2 values with increased disease severity (control T2 42.4ms vs 49.9ms, 59.7ms 92.0ms for mild/moderate/severe patient groups)(Hiba et al., 2012). Thigh muscle T2 was also found to be greater in children with active juvenile dermatomyositis than children with inactive disease or healthy children (~85ms vs ~80ms,  $p<0.05$ ). No method to account for the presence of fatty infiltration was employed, although the significant difference between patients with active and inactive disease is notable (Maillard et al., 2004a). More recently, T2 has been quantified in using T2 mapping with fat suppression in a very large cross-sectional study of boys with Duchenne muscular dystrophy – this showed increase “water-T2” in lower limb muscles which reduced with increasing age as fatty infiltration increased (Forbes et al., 2014).

Two studies have examined the magnetisation transfer ratio of muscle in neuromuscular diseases: an early study in patients with limb girdle muscular dystrophies (McDaniel et al., 1999) and a pilot study performed at our centre in patents with CMT1A and CIDP (Sinclair et al., 2012a). In both these studies MTR was reduced in muscles appearing normal on qualitative imaging, suggesting an early disease marker. Quantitative MT may be performed to model magnetisation transfer effects and derive the underlying parameters such as the restricted proton pool fraction (Sinclair et al., 2010), but these methods take significantly longer than methods to

estimate the magnetisation transfer ratio. Overall, both T2-relaxometry and magnetisation transfer imaging show promise as tools to probe abnormal muscle water distribution, and merit further study.

## **4.2.3 Correlation between MRI parameters**

### **4.2.3.1 Quantitative vs qualitative**

In DMD, Kim and colleagues showed strong correlation between T1w fat infiltration score and T2 time within gluteus maximus ( $p < 0.001$ ) (Kim et al., 2010a). Willis and colleagues showed overlapping sequential ranges of fat fraction corresponding to qualitative Mercuri T1w grades (Willis et al., 2014). In DMD, Wokke and colleagues showed strong correlation between Dixon method measured fat fraction and T1w qualitative assessment in patients with DMD, with a tendency of the qualitative assessment to over-estimate fat fraction (Wokke et al., 2013).

### **4.2.3.2 Quantitative vs quantitative**

A number of studies have compared a variety of quantitative MRI parameters and have without exception shown strong correlation. Triplett and colleagues showed strong positive correlations between muscle T1, T2, chemical-shift fat fraction and MRS fat fraction in a mixed group of DMD and Bethlem myopathy patients and controls (Triplett et al., 2014). Three fat estimation methods (2-point Dixon, T2 mapping, steady-state free precession) were highly correlated in oculopharyngeal muscular dystrophy (Gloor et al., 2011). T2 measures and 2-point Dixon showed fewer outliers in the correlations so were thought to be superior. Overall, it is clear that there is significant interdependence of T1 time, T2 time related to tissue fat fraction.

## **4.2.4 Correlation with clinical parameters**

### **4.2.4.1 Strength**

Prior to the commencement of this study, there were few publications correlating MRI measurements with patient strength, but such studies have recently exploded in number. Some qualitative studies have shown a correlation between muscle strength and T1w MRI appearance, for example in myotonic dystrophy type 1 (Coté et al., 2011). A negative correlation between muscle MRC strength grade and muscle MRI fat fraction has been shown in FSHD (Kan et al., 2009), and between myometry measurement and muscle fat fraction in myotonic dystrophy (Hiba et al., 2012) and DMD (Wokke et al., 2014). The best correlation however has been found with measures of residual muscle volume based on automatic T1w segmentation for both DM1 (Hiba et al., 2012) and SMA (Sproule et al., 2011a). Excellent correlation was also seen between thigh muscle area determined by Dixon sequence and both knee

flexion and extension in FSHD (Janssen et al., 2014). Other authors have failed to show correlation, for example no correlation was seen between muscle fat infiltration and manual muscle grading or hand held myometry in DMD (Wren et al., 2008), and no correlation between isokinetic dynamometry and muscle T1 and T2 values in patients with previous polio (Tollbäck et al., 1996). The lack of correlation may be due to inaccuracy in MRI or strength measurement, or difference in disease pathogenesis.

Correlation was found between muscle T2 times and MMT scores in juvenile dermatomyositis (Maillard et al., 2004a), between muscle T2 and maximal isometric strength in MND (Bryan et al., 1998) and between MTR and MRC grade in CMT1A (Sinclair et al., 2012a). None of these studies utilised fat separation or suppression methods, so correlation may have been due to changes in both water and fat distribution, as both affect MTR and T2 measurements. These studies have been limited in scope in terms of the number of muscles assessed, the separation of the fat and water effects, including muscle size as a covariate, and the method of strength measurement, so a more comprehensive analysis is warranted.

#### **4.2.4.2 Functional measures**

Quantitative MRI has been correlated with functional scales and tests in a number of neuromuscular diseases. Many of these are in DMD; for example in DMD there is strong correlation between gluteus maximus T2 and the clinical function score ( $p < 0.001$ ) and the timed Gower score ( $p < 0.05$ ) (Kim et al., 2010a). Other authors have shown thigh fat fraction correlates with the Brooke functional scale (Wren et al., 2008) and calf lipid fraction with 6MWT and 10m walk (Torriani et al., 2012). Timed 6MWT has also been shown to negatively correlate with hamstring fat fraction in LGMD2I (Willis et al., 2014).

It is of note that all of these measures are sensitive to either direct or indirect measures of muscle fat infiltration, which would therefore appear to be a valid measure of disease severity. For some it seems likely that the muscle chosen is able to demonstrate the spectrum of disease severity and therefore correlate with a functional measure, rather than necessarily be the primary determinant for that functional measure. For example, neither calf nor hamstring muscle fat fraction is the prime determinant of a six minute walk time. One study correlated muscle volume with the Hammersmith motor scale in spinal muscular atrophy (Sproule et al., 2011a). It is likely that both muscle atrophy and intramuscular fat infiltration are determinants of muscle weakness and disability, although their relative contribution may differ between patients and diseases. The difficulty with muscle size measurements, whether volume or cross-sectional area, is that as shown in chapter 2 these show marked intra-subject variation, though longitudinal measurements may be highly relevant.

#### **4.2.4.3 Neurophysiology**

Few studies have directly compared quantitative MRI with neurophysiology parameters in neuromuscular diseases. Neurophysiology is not central to the diagnosis of primary muscle disorders. One early study showed congruence between EMG and different imaging modalities including MRI in patients with a range of myopathic conditions (Alanen et al., 1994). In neurogenic disorders, Jonas and colleagues examined the correlations between T1w and STIR MRI sequence and EMG in tibialis anterior muscle of 20 patients with axonal polyneuropathy (Jonas et al., 2000). The measured signal intensity on the STIR images showed excellent correlation ( $\rho=0.81$ ,  $p<0.001$ ) with the degree of pathological spontaneous activity, a marker of acute denervation. The T1w signal intensity showed correlation with the motor unit amplitude, a marker of axonal loss, though this correlation was less strong, particularly if the values from healthy subjects were excluded. Deroide and colleagues examined patients with a focal lower limb neuropathy using T1w and STIR MRI sequences and EMG spontaneous activity and interference pattern (Deroide et al., 2015). Whilst both EMG and MRI parameters correlated with patient strength, there were no positive correlations between EMG and MRI parameters, perhaps limited by the wide range of muscles included.

In animal studies, longitudinal assessment is possible in controlled conditions, and a close relationship between spontaneous activity on EMG and T2 prolongation on MRI is observed, associated with enlargement in the capillary bed on muscle histology (Wessig et al., 2004). In spinal muscular atrophy, lower limb muscle volume measurements showed strong correlations ( $p<0.01$ ) with both motor unit number estimation and compound motor unit action potential amplitude (Sproule et al., 2011a). One early study in ALS was able to show correlation between muscle T2 CMAP amplitude both at baseline and in terms of longitudinal change in these parameters over 4 months (Bryan et al., 1998). No studies have compared quantitative sequences with fat fraction with neurophysiological parameters.

#### **4.2.4.4 Pathology**

Gaeta and colleagues (Gaeta et al., 2011) compared MRI measured fat fraction using a dual-echo dual flip-angle fat-water separation method, with measured fat fraction in a muscle biopsy in 27 consecutive patients with a range of neuromuscular diseases. Biopsy site was marked during the MRI protocol to ensure MRI measurement of as far as possible occurred of the tissue subsequently biopsied. There was extremely close correlation between the two fat estimates with Pearson correlation coefficient of 0.995, and Bland-Altman 95% limits of agreement (-2.9%; 2.3%). There are no other studies which have directly compared biopsy data with quantitative MRI methods, but a correlation between extensor digitorum muscle histology fat score and qualitative MRI

Mercuri grade has been shown in patients with Duchenne muscular dystrophy (Kinali et al., 2011), and correlation with MRI STIR abnormality and muscle inflammation (CD8+ T-cells) on biopsy in patients with FSHD (Frisullo et al., 2011).

#### **4.2.5 Summary**

There is a growing literature on quantitative muscle MRI in neuromuscular diseases in cross-sectional studies. Collectively these have shown that MRI can quantify intramuscular fat accumulation through a variety of methods and generally correlations with both patient strength and function are found, demonstrating validity of these outcome measures. Studies of abnormal water distribution are less numerous; the confounding effect of intramuscular fat on these measures needs to be considered. Qualitative studies in CMT1A and IBM suggest that quantitative methods to measure both acute and chronic pathology should be relevant to these diseases, and that lower limb muscles are an appropriate anatomical target.

#### **4.3 *Cross-sectional data aims***

The aims of the second phase of this thesis are to:

1. Use MRI measures of intramuscular fat and muscle size to quantify chronic pathology in CMT1A and IBM patients versus their matched control groups.
2. Assess for early pathological abnormalities by analysing T2 and MTR in muscles without significant intramuscular fat accumulation.
3. Demonstrate criterion validity of quantitative MRI outcome measures through correlation with strength and functional measures.

## 4.4 Results

### 4.4.1 Subjects

Age, gender, height and weight showed no significant differences between the two patient groups and their respective matched controls (Table 4-1).

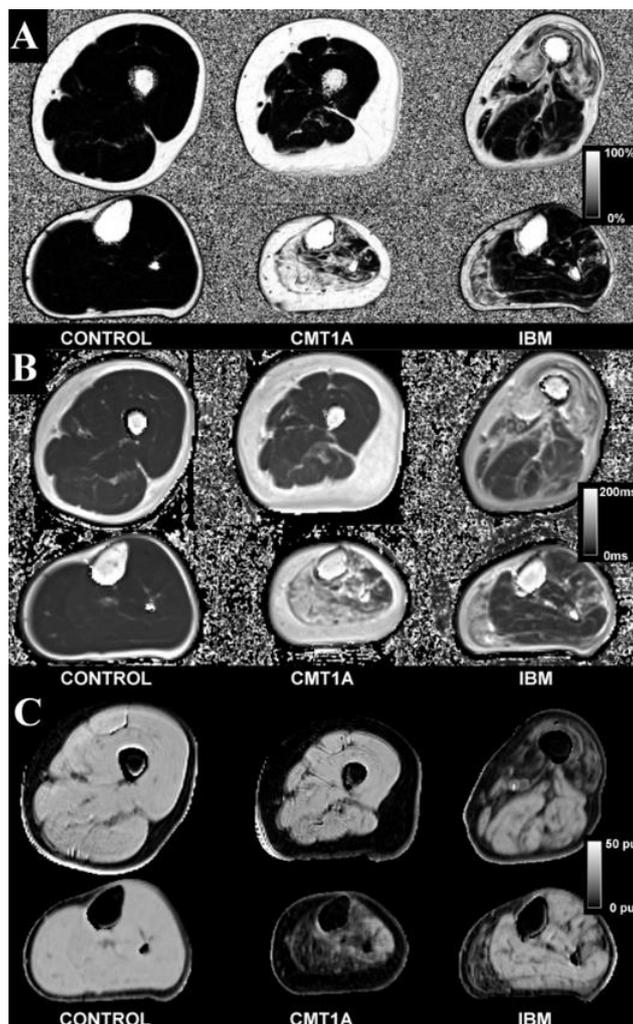
Table 4-1: Subject demographics

Measure	CMT	Control (CMT)	p	IBM	Control (IBM)	p
<b>Gender</b>	9F, 11M	9F, 11M	1	4F, 16M	8F, 12M	0.17
<b>Age</b>	42.8 ± 13.9	45.8 ± 14.2	0.51	66.7 ± 8.9	61.8 ± 10.3	0.12
<b>Height</b>	167 ± 12	171 ± 11	0.34	175 ± 8	177 ± 10	0.28
<b>Weight</b>	70 ± 16	75 ± 19	0.44	84 ± 16	77 ± 19	0.22
<b>BMI</b>	25.1 ± 4.6	25.4 ± 4.9	0.84	27.4 ± 4.0	26.0 ± 4.6	0.30

### 4.4.2 MRI data

Sample parameter maps for fat fraction, T1, T2 and MTR for each group are shown in Figure 4-1. Summary MRI measures at thigh and calf levels in patient and matched control groups are given in Table 4-2.

Figure 4-1: Sample images



Sample axial images of left lower limb at mid-thigh and mid-calf level. A) Fat fraction map of a subject from each group. B) T2 map of a subject from each group. C) MTR map of a subject from each group

Table 4-2: MRI data summary results at thigh and calf level

Measure	CMT	Control (CMT)	p	IBM	Control (IBM)	p
<b>Quantitative MRI parameters at thigh level</b>						
FF (%)	3.7 ± 6.8	1.7 ± 1.3	>0.1	26.6 ± 15.5	2.0 ± 1.2	<0.0001
T1 (ms)	1212 ± 131	1266 ± 48	>0.1	916 ± 189	1264 ± 50	<0.0001
T2 (ms)	45.7 ± 9.1	41.8 ± 3.1	0.08	83.5 ± 17.7	43.1 ± 2.5	<0.0001
MTR	30.8 ± 3.0	32.0 ± 0.8	0.09	22.9 ± 5.0	31.6 ± 0.7	<0.0001
FF whole (%)	5.8 ± 8.5	3.3 ± 1.8	>0.1	27.5 ± 15.7	4.0 ± 1.6	<0.0001
CSA (cm <sup>2</sup> )	201 ± 53	219 ± 56	>0.1	169 ± 46	212 ± 51	0.009
<b>Qualitative MRI at thigh level</b>						
Mercuri (0-5)	0.6 ± 0.8	0.3 ± 0.3	>0.1	2.7 ± 1.0	0.4 ± 0.3	<0.0001
STIR (0-2)	0.2 ± 0.3	0.0 ± 0.1	>0.1	0.8 ± 0.4	0.1 ± 0.1	<0.0001
<b>MRI parameters at calf level</b>						
FF (%)	15.8 ± 25.5	1.6 ± 1.0	0.02	18.0 ± 13.2	2.0 ± 0.9	<0.0001
T1 (ms)	1080 ± 314	1247 ± 52	<0.05	999 ± 202	1249 ± 54	<0.001
T2 (ms)	59.7 ± 29.3	40.2 ± 3.6	0.005	71.1 ± 21.4	41.9 ± 3.8	<0.0001
MTR	26.1 ± 9.0	32.1 ± 1.0	0.007	23.7 ± 5.7	31.6 ± 0.9	<0.0001
FF whole (%)	15.5 ± 24.0	2.7 ± 1.5	0.03	19.2 ± 13.7	3.5 ± 1.3	<0.0001
CSA (cm <sup>2</sup> )	100 ± 26	123 ± 27	0.01	112 ± 26	120 ± 31	>0.1
<b>Qualitative MRI at calf level</b>						
Mercuri (0-5)	1.5 ± 1.5***	0.2 ± 0.3	<0.001	1.9 ± 1.0***	0.3 ± 0.3	<0.001
STIR (0-2)	0.4 ± 0.3	0.3 ± 0.3	>0.1	0.7 ± 0.4**	0.4 ± 0.3	<0.01

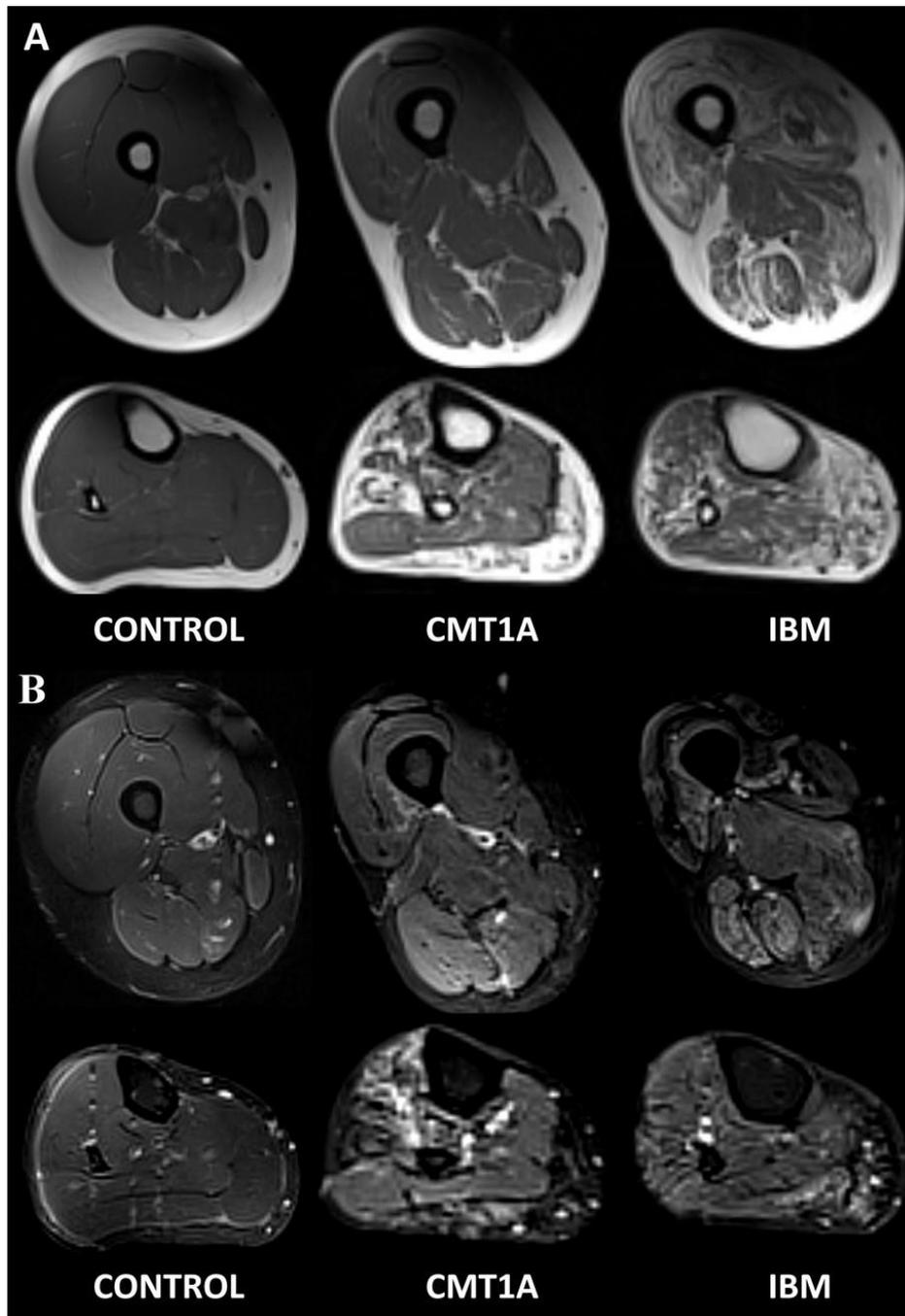
Data are presented mean ± standard deviation by subject. Significant differences to matched controls are indicated: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . CMT: Charcot-Marie-Tooth disease; IBM: inclusion body myositis; FF: fat fraction; FFw: whole muscle fat fraction; MTR: magnetisation transfer ratio; STIR: short-tau-inversion-recovery

#### 4.4.3 Measures of chronic fatty atrophy

Whilst all MRI sequences utilised in this study are sensitive to intramuscular fat, the qualitative T1-weighted sequence, and quantitative T1-mapping and three-point Dixon sequences were included as the primary measures of intramuscular fat accumulation.

#### 4.4.3.1 Qualitative T1w sequences

Figure 4-2: Qualitative imaging and Regions of Interest



Sample axial T1-weighted (A) and STIR (B) images of right mid-thigh and mid-calf. Within muscles fat infiltration appears hyperintense on T1w images, whilst muscle oedema appears hyperintense on STIR images. Left) The healthy volunteer has low signal intensity on both sequences, however field homogeneity is evident which hinders quantitative analysis of T1w and STIR sequences Middle) In this typical CMT1A patient thigh muscle returns normal signal intensity, but there is fat infiltration at calf level, most marked in tibialis anterior, peroneus longus and both heads of gastrocnemius, and muscle oedema in tibialis anterior. Right) In this typical IBM patient fat infiltration and muscle oedema is present at both thigh and calf level, most markedly in quadriceps muscle where there is also muscle atrophy.

The frequency of Mercuri grading of T1w images within thigh and calf muscles are given in Table 4-3. In the healthy controls, no muscles were graded Mercuri 2b or higher at thigh or calf level. Whilst the overall median grades for both T1w images were not significantly different in the CMT1A group as a whole, some muscles were graded 2b (4.5%) and 3 (2.0%). These were from just 3/20 patients, whose T1w thigh scans could therefore be considered qualitatively abnormal. By contrast at calf level the full spectrum of Mercuri grades were assigned to CMT1A patients' muscles, with only 39.6% of muscles grade 0. This is consistent with the length dependent distribution of weakness observed clinically in CMT1A patients. The IBM patients showed more marked T1w abnormalities with Mercuri grade 3 most commonly assigned at thigh level, with slightly lesser grades in the calf, but over half of muscles graded 2a or higher. On qualitative assessment, the distribution of chronic pathology in IBM is therefore more widespread, but overall with greater thigh than calf involvement.

*Table 4-3: Distribution of qualitative Mercuri grading by subject group*

Thigh		Mercuri							Total	
		0	1	2a	2b	3	4	Art.		
Diagnosis	CMT1A	Count	219	118	26	18	8	0	11	400
		% within Diagnosis	54.8%	29.5%	6.5%	4.5%	2.0%	0.0%	2.8%	100%
	IBM	Count	36	80	68	57	116	43	0	400
		% within Diagnosis	9.0%	20.0%	17.0%	14.3%	29.0%	10.8%	0.0%	100%
	Volunteer	Count	423	137	15	0	0	0	5	580
		% within Diagnosis	72.9%	23.6%	2.6%	0.0%	0.0%	0.0%	0.9%	100%
Total		Count	678	335	109	75	124	43	16	1380
		% within Diagnosis	49.1%	24.3%	7.9%	5.4%	9.0%	3.1%	1.2%	100.0%

Calf		Mercuri							Total	
		0	1	2a	2b	3	4	Art.		
Diagnosis	CMT1A	Count	95	52	34	14	24	21	0	240
		% within Diagnosis	39.6%	21.7%	14.2%	5.8%	10.0%	8.8%	0.0%	100%
	IBM	Count	55	64	40	31	32	18	0	240
		% within Diagnosis	22.9%	26.7%	16.7%	12.9%	13.3%	7.5%	0.0%	100%
	Volunteer	Count	279	65	4	0	0	0	0	348
		% within Diagnosis	80.2%	18.7%	1.1%	0.0%	0.0%	0.0%	0.0%	100%
Total		Count	429	181	78	45	56	39	0	828
		% within Diagnosis	51.8%	21.9%	9.4%	5.4%	6.8%	4.7%	0.0%	100.0%

*Art.: Artefact. Description of Mercuri grades are in Figure 3-11.*

#### 4.4.3.2 Overall quantitative MRI results

Summary quantitative MRI data on a per subject basis are in Table 4-2. The direction of differences seen in patients was consistent with that expected with intramuscular fat accumulation: increased FF and T2 and reduced T1 and MTR. In IBM patients all quantitative MRI parameters were significantly different ( $p < 0.001$ ) from controls at thigh and calf level. Total cross-sectional area was reduced at thigh ( $p < 0.01$ ) but not calf

level. In CMT patients, with the length dependent clinical phenotype, and consistent with the qualitative findings, there were significant ( $p<0.05$ ) differences in all quantitative measure at calf but not thigh level. Total calf cross-sectional area was also reduced. These data demonstrate quantifiable abnormalities consistent with chronic fatty atrophy of lower limb muscles.

#### 4.4.3.3 Distribution of MRI abnormalities

Muscle-specific results in both patient groups and their matched control groups for cross-sectional area are given in Table 4-4 and quantitative MRI parameters in Table 4-5 (CMT1A) and Table 4-6 (IBM).

Table 4-4: Cross-sectional area by muscle in patients and matched controls

Group	CMT	Control (CMT)	IBM	Control (IBM)
<b>Thigh</b>				
<i>Rectus Femoris</i>	571±275	735±304	405±304	618±280
<i>Vastus Medialis</i>	2075±805	2206±786	1251±458	1898±903
<i>Vastus Intermedius</i>	2200±675	2817±775	1773±601	2633±971
<i>Vastus Lateralis</i>	2928±876	2966±882	1969±681	2448±857
<i>Semimembranosus</i>	1218±489	1482±511	1467±894	1452±563
<i>Semitendinosus</i>	971±371	1155±458	1034±424	1058±429
<i>Biceps Femoris</i>	2054±619	2251±739	2045±894	2213±893
<i>Adductor Magnus</i>	2547±1079	2378±813	2337±1307	2291±931
<i>Sartorius</i>	576±144	550±186	491±204	538±226
<i>Gracilis</i>	565±195	559±213	360±180	561±229
<b>Total</b>	15700±10500	17100±10750	13130±9660	15710±10470
<b>Calf</b>				
<i>Tibialis Anterior</i>	1117±322	1643±429	1479±520	1524±684
<i>Peroneus Longus</i>	855±286	868±260	802±348	839±399
<i>Lateral Gastrocnemius</i>	1026±298	1310±530	1097±544	1097±564
<i>Medial Gastrocnemius</i>	1819±635	2234±685	1369±657	1848±880
<i>Soleus</i>	2993±1023	3410±718	3229±1285	3098±1343
<i>Tibialis Posterior</i>	425±122	644±217	678±246	558±280
<b>Total</b>	8238±5946	10110±6354	8652±6528	8964±6798

Results are mean ± s.d. in mm<sup>2</sup>

Table 4-5: Muscle-specific MRI parameters in CMT1A and matched controls

Expressed as mean ± s.d (n); FFw: fat fraction from whole muscle ROI, other values from small ROI

Muscle	CMT1A					Volunteer (CMT1A)				
	FF (%)	T1 (ms)	T2 (ms)	MTR (p.u.)	FFw (%)	FF (%)	T1 (ms)	T2 (ms)	MTR (p.u.)	FFw (%)
Right Rectus Femoris	3.6 ± 12.1 (20)	511 ± 0 (1)	46.8 ± 13.7 (16)	28.9 ± 9.3 (5)	6.3 ± 13 (20)	0.7 ± 1 (20)	1158 ± 0 (1)	40.8 ± 5.5 (17)	33.3 ± 1.9 (3)	2.7 ± 1.9 (19)
Right Vastus Medialis	2 ± 2.9 (20)	1270 ± 170 (7)	43.8 ± 7.8 (20)	31.6 ± 2.2 (10)	4.4 ± 8.1 (20)	1 ± 0.8 (20)	1343 ± 80 (8)	40.9 ± 3.9 (19)	32.6 ± 1.9 (8)	2.2 ± 1.1 (19)
Right Vastus Intermedius	3 ± 5.8 (20)	1232 ± 175 (9)	44.6 ± 11.5 (20)	30.3 ± 4.8 (14)	4.3 ± 10.4 (20)	1 ± 1.1 (20)	1376 ± 110 (10)	38.4 ± 2.8 (20)	33.2 ± 1.8 (14)	2 ± 1.6 (19)
Right Vastus Lateralis	5.4 ± 19.1 (20)	1293 ± 262 (16)	49.6 ± 26.2 (19)	31 ± 6.6 (20)	6.2 ± 14 (20)	1.2 ± 0.9 (20)	1368 ± 54 (11)	42.2 ± 4.8 (18)	32.5 ± 1.3 (19)	2.6 ± 1.3 (19)
Right Semimembranosus	3.5 ± 4.5 (20)	1175 ± 145 (18)	48.5 ± 8.8 (19)	30.5 ± 2.9 (20)	6.7 ± 5.5 (20)	2.5 ± 3.1 (20)	1212 ± 116 (17)	44.9 ± 6.4 (20)	31.4 ± 1.6 (18)	4.6 ± 2.9 (19)
Right Semitendinosus	2.3 ± 2.7 (20)	1219 ± 141 (17)	47.4 ± 8.3 (20)	31.2 ± 2.8 (20)	6.3 ± 5.3 (20)	1.1 ± 0.7 (20)	1306 ± 86 (18)	43.3 ± 4 (20)	32.1 ± 1.1 (18)	3.6 ± 2.7 (19)
Right Biceps Femoris	2.7 ± 3.5 (20)	1253 ± 130 (17)	47.8 ± 7.4 (19)	30.5 ± 2.6 (20)	4.6 ± 4.7 (20)	2 ± 1.4 (20)	1283 ± 86 (18)	46.5 ± 5 (18)	31.2 ± 1.3 (18)	3.7 ± 2.4 (19)
Right Adductor Magnus	2.3 ± 4 (20)	1303 ± 116 (17)	45 ± 7.3 (20)	31.9 ± 2.2 (20)	5.1 ± 7.3 (20)	2 ± 3 (20)	1308 ± 109 (16)	43 ± 5.6 (20)	32.1 ± 1.6 (18)	3.8 ± 3.4 (19)
Right Sartorius	5.9 ± 10.6 (20)	1149 ± 208 (13)	48.3 ± 11.1 (20)	27.7 ± 5.4 (13)	9.5 ± 8.6 (20)	3.6 ± 3.8 (20)	1245 ± 101 (12)	43.6 ± 6.1 (19)	31.9 ± 1.8 (8)	5.9 ± 2.8 (19)
Right Gracilis	3.1 ± 4.6 (19)	1188 ± 146 (16)	44 ± 7.5 (20)	31.8 ± 3.2 (19)	6.8 ± 5.7 (20)	1.8 ± 1.7 (20)	1285 ± 85 (16)	39.5 ± 3.5 (20)	32.8 ± 1.4 (17)	4.1 ± 2.3 (18)
Left Rectus Femoris	2.7 ± 8.2 (20)	1257 ± 258 (12)	45.4 ± 17.1 (20)	31.6 ± 4.6 (16)	6 ± 11.4 (20)	1.1 ± 1.6 (20)	1384 ± 76 (8)	38.4 ± 4.8 (18)	32.8 ± 1.9 (18)	3 ± 2.3 (19)
Left Vastus Medialis	2.7 ± 5.4 (20)	1284 ± 151 (15)	44.3 ± 9.2 (20)	31.5 ± 3.4 (19)	5.3 ± 9.3 (20)	1.2 ± 1.3 (20)	1359 ± 45 (12)	42 ± 4.7 (20)	32.3 ± 1.2 (18)	2.7 ± 1.4 (19)
Left Vastus Intermedius	8.2 ± 21.7 (20)	1210 ± 322 (15)	51.8 ± 29.6 (20)	29 ± 7.7 (19)	6.1 ± 12.6 (20)	1.3 ± 1 (20)	1331 ± 65 (13)	42.3 ± 5.5 (19)	31.7 ± 1.4 (19)	2 ± 1.3 (19)
Left Vastus Lateralis	4.8 ± 15.1 (20)	1213 ± 194 (19)	44.5 ± 18.7 (20)	30.8 ± 5.6 (19)	7.2 ± 15.1 (20)	1.6 ± 1.2 (20)	1286 ± 56 (17)	40.6 ± 3 (19)	32.1 ± 1.2 (19)	3.4 ± 1.7 (19)
Left Semimembranosus	3 ± 1.8 (20)	1077 ± 81 (16)	43.5 ± 4.4 (20)	29.6 ± 2.4 (16)	6.9 ± 5.5 (20)	2.3 ± 1.8 (19)	1130 ± 66 (18)	42 ± 5.1 (20)	30.9 ± 1.1 (17)	4.2 ± 2.6 (17)
Left Semitendinosus	2.6 ± 3.1 (20)	1116 ± 87 (18)	42.7 ± 5.6 (20)	31.1 ± 2.6 (18)	6.4 ± 5.4 (20)	1.4 ± 1.2 (19)	1168 ± 60 (19)	40.8 ± 4.3 (20)	31.9 ± 1 (18)	4.1 ± 2.3 (17)
Left Biceps Femoris	3 ± 4.2 (20)	1162 ± 123 (19)	44.8 ± 7.4 (20)	30.7 ± 3.2 (19)	6.1 ± 5.7 (19)	2.6 ± 2.2 (20)	1201 ± 72 (18)	43.5 ± 6.4 (20)	31.8 ± 1.4 (19)	4.8 ± 2.8 (19)
Left Adductor Magnus	2 ± 2.2 (20)	1194 ± 112 (17)	41.9 ± 4.2 (20)	30.7 ± 2.1 (18)	6.2 ± 8.9 (19)	2 ± 2.3 (20)	1246 ± 49 (15)	42.4 ± 5.2 (20)	31.6 ± 1.3 (18)	3.9 ± 3.3 (19)
Left Sartorius	5.5 ± 9.9 (20)	1210 ± 245 (12)	47.4 ± 13.2 (20)	30.2 ± 4.6 (17)	9.1 ± 9.1 (20)	2.7 ± 2.9 (20)	1266 ± 77 (11)	42 ± 6 (20)	31.8 ± 2.3 (16)	6.7 ± 3.5 (19)
Left Gracilis	5 ± 11.5 (20)	1161 ± 198 (13)	43.7 ± 13.4 (20)	30.4 ± 4.2 (18)	7.5 ± 9.7 (20)	1.8 ± 1.6 (20)	1254 ± 74 (17)	38.1 ± 3.8 (20)	32.4 ± 1.4 (18)	5 ± 2 (19)
Right Tibialis Anterior	12.6 ± 23.6 (20)	845 ± 272 (10)	57.2 ± 29.7 (20)	23.4 ± 8.4 (16)	14 ± 21.2 (20)	0.7 ± 0.7 (19)	1133 ± 115 (9)	36.9 ± 2.3 (20)	30.5 ± 1.4 (17)	1.5 ± 0.9 (19)
Right Peroneus Longus	21.9 ± 30.5 (20)	895 ± 355 (14)	66.4 ± 36.1 (20)	24.2 ± 10.9 (20)	24.4 ± 28.6 (20)	1.6 ± 1.5 (19)	1218 ± 104 (14)	40.6 ± 5.1 (19)	31.9 ± 1.6 (18)	3.1 ± 2.4 (19)
Right Lateral Gastrocnemius	15.1 ± 29 (20)	1081 ± 383 (18)	60.2 ± 34 (20)	25.8 ± 10.2 (20)	15.8 ± 27.3 (20)	1.3 ± 0.9 (19)	1238 ± 69 (15)	40.4 ± 4.7 (20)	32.1 ± 0.9 (17)	2.5 ± 1.5 (19)
Right Medial Gastrocnemius	18.2 ± 29.1 (20)	1021 ± 386 (16)	61.5 ± 30.8 (20)	25.3 ± 10.2 (19)	18.6 ± 28.1 (20)	1.5 ± 0.9 (18)	1220 ± 44 (17)	40.2 ± 3.5 (20)	32.7 ± 1.3 (17)	2.5 ± 1.2 (18)
Right Soleus	13 ± 25.8 (20)	1086 ± 343 (16)	56.4 ± 29.6 (20)	27.6 ± 9.4 (19)	12.6 ± 23.7 (20)	2.7 ± 2.9 (19)	1231 ± 68 (13)	43 ± 6.5 (20)	32 ± 1.6 (18)	3 ± 2.7 (19)
Right Tibialis Posterior	12.4 ± 25.9 (20)	1052 ± 353 (13)	56.5 ± 34.2 (19)	27.7 ± 9.5 (20)	12.4 ± 22.6 (20)	1 ± 0.9 (19)	1278 ± 86 (12)	38.9 ± 2.8 (20)	32.9 ± 1.1 (18)	1.5 ± 1 (19)
Left Tibialis Anterior	16.8 ± 28.4 (19)	1056 ± 343 (18)	64.5 ± 35.5 (19)	25.4 ± 9.4 (20)	15.7 ± 22.2 (19)	0.9 ± 0.7 (19)	1298 ± 83 (17)	40.3 ± 5.9 (20)	32 ± 1.4 (18)	2 ± 1 (19)
Left Peroneus Longus	18.9 ± 30 (20)	1057 ± 325 (18)	61.1 ± 32.5 (19)	25.8 ± 10.4 (20)	22.4 ± 27.8 (20)	2 ± 1.6 (19)	1212 ± 103 (17)	41.1 ± 5.8 (20)	32 ± 1.8 (19)	3.7 ± 2.3 (19)
Left Lateral Gastrocnemius	16.8 ± 29.3 (20)	1087 ± 395 (17)	60.1 ± 34 (19)	26 ± 10.6 (20)	17.8 ± 27.7 (20)	1.8 ± 1.6 (19)	1274 ± 78 (17)	39.4 ± 3.9 (20)	32.1 ± 1.1 (18)	2.8 ± 1.7 (18)
Left Medial Gastrocnemius	20 ± 32.5 (20)	1024 ± 389 (17)	64 ± 33.8 (19)	25.8 ± 10.8 (20)	19.7 ± 29.8 (19)	1.9 ± 1.3 (19)	1248 ± 49 (17)	39.7 ± 3.6 (20)	32.4 ± 1 (18)	3.9 ± 1.6 (19)
Left Soleus	12.3 ± 23.3 (20)	1136 ± 324 (18)	55.6 ± 27 (19)	28 ± 8.7 (20)	13.3 ± 22.9 (19)	2.4 ± 1.8 (19)	1254 ± 75 (17)	41.5 ± 4.2 (20)	32 ± 1.4 (19)	2.9 ± 2.3 (17)
Left Tibialis Posterior	11.5 ± 21.8 (20)	1143 ± 324 (18)	58.6 ± 30.2 (19)	27.1 ± 8.5 (20)	14.1 ± 22 (20)	1.4 ± 0.8 (19)	1295 ± 80 (17)	39.5 ± 2.5 (20)	31.9 ± 1.6 (19)	2.3 ± 1.3 (19)

Table 4-6: Muscle-specific MRI parameters in IBM and matched controls

Expressed as mean ± s.d (n); FFw: fat fraction from whole muscle ROI, other values from small ROI.

Muscle	IBM					Volunteer (IBM)				
	FF (%)	T1 (ms)	T2 (ms)	MTR (p.u.)	FFw (%)	FF (%)	T1 (ms)	T2 (ms)	MTR (p.u.)	FFw (%)
Right Rectus Femoris	19 ± 16.8 (20)	881 ± 159 (8)	81.3 ± 29.1 (19)	20.6 ± 8.2 (9)	26.8 ± 17.6 (20)	0.6 ± 0.7 (19)	1201 ± 60 (2)	41.8 ± 4.5 (16)	30.7 ± 2.3 (3)	3.1 ± 1.5 (19)
Right Vastus Medialis	31.4 ± 22.7 (20)	799 ± 248 (13)	101.2 ± 27.6 (19)	16.9 ± 6.6 (9)	37.4 ± 17.5 (20)	1.4 ± 0.7 (19)	1342 ± 94 (7)	43.8 ± 4.3 (17)	31.9 ± 2.4 (5)	2.7 ± 1 (19)
Right Vastus Intermedius	37.6 ± 26.5 (20)	648 ± 224 (15)	99.4 ± 30.7 (20)	15.5 ± 9.1 (15)	33.9 ± 20.5 (20)	1.1 ± 0.9 (19)	1394 ± 69 (8)	39.3 ± 2.4 (18)	32.8 ± 1.8 (14)	2.7 ± 1.6 (19)
Right Vastus Lateralis	49.2 ± 27.8 (20)	634 ± 289 (18)	115.9 ± 31.3 (20)	12.8 ± 8.6 (16)	44.1 ± 24 (20)	1.5 ± 0.9 (19)	1366 ± 58 (13)	42.7 ± 3.7 (17)	32 ± 1.2 (18)	3.2 ± 1.1 (19)
Right Semimembranosus	19.6 ± 23.8 (20)	944 ± 307 (17)	71.2 ± 28.4 (20)	23.6 ± 8.9 (18)	23 ± 22.6 (20)	3.1 ± 2.9 (19)	1183 ± 123 (16)	47.4 ± 6.6 (18)	30.8 ± 1.6 (16)	5.7 ± 2.4 (19)
Right Semitendinosus	19.1 ± 28.3 (20)	1135 ± 352 (17)	68.6 ± 29.1 (20)	26.4 ± 7.5 (17)	23.5 ± 25.8 (20)	1.1 ± 0.6 (19)	1303 ± 72 (17)	43.5 ± 3.5 (18)	32.1 ± 0.8 (17)	4.4 ± 2.5 (19)
Right Biceps Femoris	6.6 ± 10.2 (20)	1257 ± 180 (17)	59.7 ± 17.2 (19)	28.6 ± 4.3 (18)	9.5 ± 8.4 (20)	2.6 ± 1.2 (19)	1268 ± 101 (17)	48 ± 4 (17)	30.5 ± 1.1 (17)	4.1 ± 2.2 (19)
Right Adductor Magnus	34.3 ± 38.5 (20)	857 ± 505 (12)	89.1 ± 41 (19)	22.3 ± 11 (16)	34.5 ± 32.8 (20)	2.4 ± 3 (19)	1282 ± 112 (16)	44.4 ± 5.7 (18)	31.5 ± 1.7 (17)	4.6 ± 3.3 (19)
Right Sartorius	22.4 ± 24.7 (20)	977 ± 291 (12)	74.4 ± 24.3 (20)	24.6 ± 5.5 (11)	26.7 ± 22.6 (20)	3.5 ± 3.4 (19)	1248 ± 133 (10)	44.1 ± 5.8 (16)	30.8 ± 1.5 (10)	6.7 ± 2.8 (19)
Right Gracilis	27.4 ± 31.1 (20)	973 ± 350 (17)	71.8 ± 30.1 (20)	24 ± 8.6 (15)	29.1 ± 26.4 (20)	1.6 ± 1.2 (19)	1306 ± 81 (16)	40.4 ± 3.5 (17)	32.9 ± 1.6 (17)	4.5 ± 2 (17)
Left Rectus Femoris	29.5 ± 25.5 (20)	736 ± 282 (15)	98.3 ± 39.6 (19)	18.7 ± 7.4 (10)	33.8 ± 22.9 (20)	1 ± 0.7 (19)	1383 ± 65 (7)	39.4 ± 5 (17)	32.6 ± 1.8 (18)	3.3 ± 1.8 (19)
Left Vastus Medialis	37.5 ± 25.5 (20)	646 ± 239 (13)	114.2 ± 39.1 (19)	16.4 ± 9.3 (13)	39.2 ± 20.7 (20)	1.4 ± 1.1 (19)	1377 ± 49 (12)	42.3 ± 4.2 (19)	32.4 ± 1.2 (18)	3.2 ± 1.2 (19)
Left Vastus Intermedius	38.4 ± 27.5 (20)	695 ± 284 (15)	119.4 ± 41.4 (18)	17.4 ± 8.1 (13)	38 ± 21.7 (20)	1.4 ± 1 (19)	1346 ± 78 (14)	43.4 ± 5 (19)	31.6 ± 1.7 (18)	2.4 ± 1.1 (19)
Left Vastus Lateralis	46.4 ± 28.5 (20)	621 ± 302 (17)	108.4 ± 32.1 (19)	14.7 ± 9.2 (14)	44.2 ± 21.1 (20)	2 ± 1 (19)	1286 ± 61 (17)	42 ± 2.1 (19)	31.7 ± 0.9 (18)	3.9 ± 1.3 (19)
Left Semimembranosus	10.9 ± 15.3 (20)	1067 ± 163 (18)	55.8 ± 15.9 (20)	27.6 ± 2.5 (16)	17.5 ± 14.5 (20)	2.8 ± 1.7 (18)	1114 ± 67 (17)	44.5 ± 5.4 (19)	29.9 ± 1.4 (15)	5.1 ± 2.5 (18)
Left Semitendinosus	16.1 ± 26.1 (20)	1031 ± 266 (18)	60.3 ± 24 (20)	28.1 ± 4.5 (15)	21 ± 23.3 (20)	1.4 ± 0.8 (19)	1167 ± 62 (18)	42.3 ± 4.6 (19)	31.4 ± 0.9 (16)	4.8 ± 2 (18)
Left Biceps Femoris	6.1 ± 8.2 (20)	1177 ± 173 (19)	54.9 ± 14.4 (20)	29.6 ± 1.7 (16)	11.4 ± 8.7 (20)	2.9 ± 2.1 (19)	1209 ± 69 (18)	44.8 ± 6.1 (19)	31.4 ± 1.2 (18)	5.3 ± 2.7 (19)
Left Adductor Magnus	33.9 ± 36.4 (20)	828 ± 414 (16)	87.5 ± 41.5 (19)	20.3 ± 11.4 (14)	29 ± 28.5 (20)	2.5 ± 2.3 (19)	1224 ± 44 (15)	44 ± 4.9 (19)	31.1 ± 1.2 (17)	4.7 ± 3.2 (19)
Left Sartorius	25.8 ± 27.4 (20)	995 ± 376 (14)	81.5 ± 33.3 (20)	24.7 ± 7 (13)	29.9 ± 25.6 (20)	3.4 ± 3 (19)	1263 ± 73 (10)	43.7 ± 6.2 (19)	31.1 ± 1.4 (15)	7.7 ± 3.3 (19)
Left Gracilis	21.2 ± 23.2 (20)	1046 ± 323 (14)	63.4 ± 22.1 (20)	26.6 ± 5.8 (12)	24.5 ± 19.3 (20)	1.7 ± 1.4 (19)	1250 ± 82 (16)	39.4 ± 3.8 (19)	31.9 ± 1.6 (16)	5 ± 2 (19)
Right Tibialis Anterior	7.1 ± 14.6 (17)	1031 ± 190 (12)	64.2 ± 35 (18)	25.3 ± 7.1 (13)	7.4 ± 10.6 (17)	1 ± 0.5 (18)	1096 ± 53 (9)	38.5 ± 2.5 (19)	29.9 ± 1.2 (15)	1.9 ± 0.9 (17)
Right Peroneus Longus	9 ± 13.4 (17)	1004 ± 244 (13)	65.4 ± 32.5 (18)	26.6 ± 6.6 (18)	12.4 ± 10.6 (17)	1.7 ± 1.5 (18)	1203 ± 117 (10)	41.5 ± 4.8 (18)	31.4 ± 1.6 (15)	4 ± 2.5 (17)
Right Lateral Gastrocnemius	20.9 ± 29 (17)	1015 ± 342 (15)	75.4 ± 37.8 (19)	23.6 ± 9.9 (18)	23.3 ± 25.5 (17)	1.6 ± 1 (18)	1248 ± 79 (15)	43.2 ± 6.6 (19)	31.4 ± 1.2 (16)	3 ± 1.4 (17)
Right Medial Gastrocnemius	53 ± 32 (17)	613 ± 337 (15)	106.4 ± 29.8 (18)	12.9 ± 9.7 (18)	50.6 ± 25.7 (17)	1.6 ± 0.8 (18)	1224 ± 49 (15)	42.3 ± 4.5 (19)	32 ± 1.7 (16)	3.2 ± 1.4 (17)
Right Soleus	9.2 ± 17.1 (17)	1161 ± 253 (12)	59.8 ± 26.7 (18)	27.5 ± 6.7 (16)	13.1 ± 15.1 (17)	3.1 ± 2.7 (18)	1210 ± 62 (13)	43.9 ± 6.2 (19)	31.5 ± 1.4 (17)	3.9 ± 2.6 (17)
Right Tibialis Posterior	3.8 ± 7.2 (17)	1216 ± 135 (10)	48.6 ± 15.6 (18)	30.8 ± 3.5 (16)	5.8 ± 9.2 (17)	1.3 ± 0.7 (18)	1283 ± 77 (14)	39.8 ± 2.2 (19)	32.6 ± 1.1 (17)	2.1 ± 0.7 (17)
Left Tibialis Anterior	8.9 ± 14.7 (18)	1201 ± 278 (12)	66.6 ± 33.7 (18)	26.2 ± 7 (17)	9.4 ± 11.4 (18)	1.2 ± 0.7 (18)	1296 ± 93 (17)	41.9 ± 5.8 (19)	31.5 ± 1.5 (16)	2.5 ± 1 (17)
Left Peroneus Longus	11 ± 12.3 (18)	1046 ± 234 (17)	64.7 ± 32.2 (19)	25.7 ± 7.7 (19)	15.2 ± 11.4 (18)	2.4 ± 2.1 (18)	1200 ± 103 (17)	42.7 ± 6 (19)	31.2 ± 1.6 (17)	4.6 ± 2.8 (17)
Left Lateral Gastrocnemius	18.5 ± 28.3 (18)	1053 ± 353 (16)	70.7 ± 34.2 (19)	22.9 ± 10.1 (19)	23.2 ± 24.5 (18)	2.1 ± 1.8 (18)	1312 ± 80 (17)	42.2 ± 5.2 (19)	31.5 ± 1.1 (16)	3.5 ± 1.6 (16)
Left Medial Gastrocnemius	54.7 ± 29.7 (18)	616 ± 290 (16)	111 ± 28.9 (19)	12.5 ± 8.5 (19)	52.9 ± 24.5 (18)	2.9 ± 2.1 (18)	1268 ± 63 (17)	42.1 ± 6.4 (19)	32 ± 0.9 (16)	4.6 ± 1.7 (17)
Left Soleus	14.1 ± 20 (18)	1078 ± 286 (15)	63.6 ± 30.3 (19)	26.4 ± 8.3 (19)	17.8 ± 18.8 (18)	3.2 ± 1.7 (18)	1246 ± 74 (17)	43.2 ± 4.9 (19)	31.6 ± 1.4 (17)	4.1 ± 2.1 (15)
Left Tibialis Posterior	6.5 ± 17.5 (18)	1315 ± 83 (14)	54.3 ± 29.8 (19)	29.1 ± 6.2 (18)	9.3 ± 14.7 (18)	1.6 ± 1 (18)	1283 ± 84 (16)	40.5 ± 2.3 (19)	31.5 ± 1.6 (17)	2.9 ± 1.4 (17)

#### **4.4.3.3.1 Disease distribution in IBM**

The significant atrophy of thigh musculature (Table 4-2) was predominantly in the quadriceps muscle, but was unexpectedly most marked in gracilis muscle, where patient cross-sectional area was 64% of their matched controls ( $p < 0.01$ ). There was no significant difference in CSA of the hamstring muscles, adductor magnus or sartorius. Fat fraction was similarly highest in the quadriceps muscle (Figure 4-3), with relative sparing of rectus femoris. The hamstrings were generally spared, especially biceps femoris, though there were many outliers. Adductor magnus showed the widest variability with the fourth lowest median FF, but highest 75<sup>th</sup> percentile. The other quantitative parameters showed the same distribution as fat fraction, due to their high inter-correlation. The same pattern of IFA was seen qualitatively (Figure 4-4).

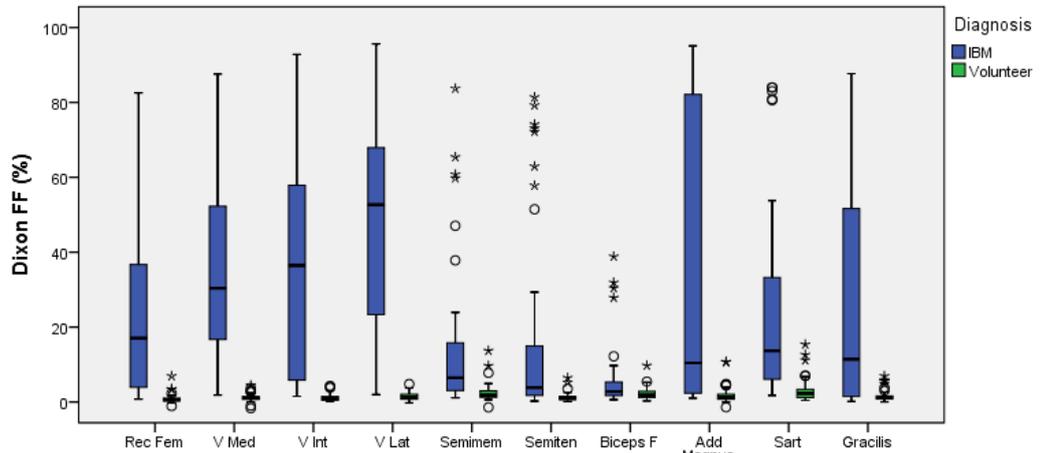
Although total calf cross-sectional area was not reduced in IBM patients, there was selective atrophy of medial gastrocnemius, also the muscle with highest fat fraction. This raises the possibility of compensatory hypertrophy in the less affected muscles. Lateral gastrocnemius had second greatest fat fraction, with relative sparing of tibialis posterior. Overall mean values were similar in right and left lower limbs, but there was significant asymmetry in some muscles of some patients. Whole muscle fat fraction was generally higher than small ROI fat fraction. This was true for all muscles in controls, but not all muscles in patients (Table 4-6).

#### **4.4.3.3.2 Disease distribution in CMT1A**

In CMT1A there was overall atrophy at the calf (Table 4-2), with all assessed muscles of reduced size compared with controls except peroneus longus (Table 4-4). Peroneus longus however had the highest mean fat fraction – suggesting some discordance between the pathological processes underlying atrophy and intramuscular fat accumulation. The pattern in individual patients however was quite variable (Figure 4-5). Some patients had severe fat infiltration of all calf muscles (blue lines in Figure 4-5) whilst some had low fat fraction in all muscles (green lines in Figure 4-5). As with IBM patients, fat fractions from whole ROI were overall slightly higher than small ROI FF, whilst T2, MTR and T1 distribution of abnormalities was the same as for FF.

Within the thigh, whilst the group statistics didn't differ significantly, in the two patients who had severe changes at the calf level, there were abnormalities at the thigh level, with relative predilection for quadriceps muscles.

Figure 4-3: Distribution of intramuscular fat accumulation in thigh of IBM patients



Bars indicate median, 25<sup>th</sup>, 50<sup>th</sup> and 75<sup>th</sup> centiles, lines: range, °: minor outlier, \*: major outlier

Figure 4-4: Qualitative and quantitative pattern of IFA in IBM

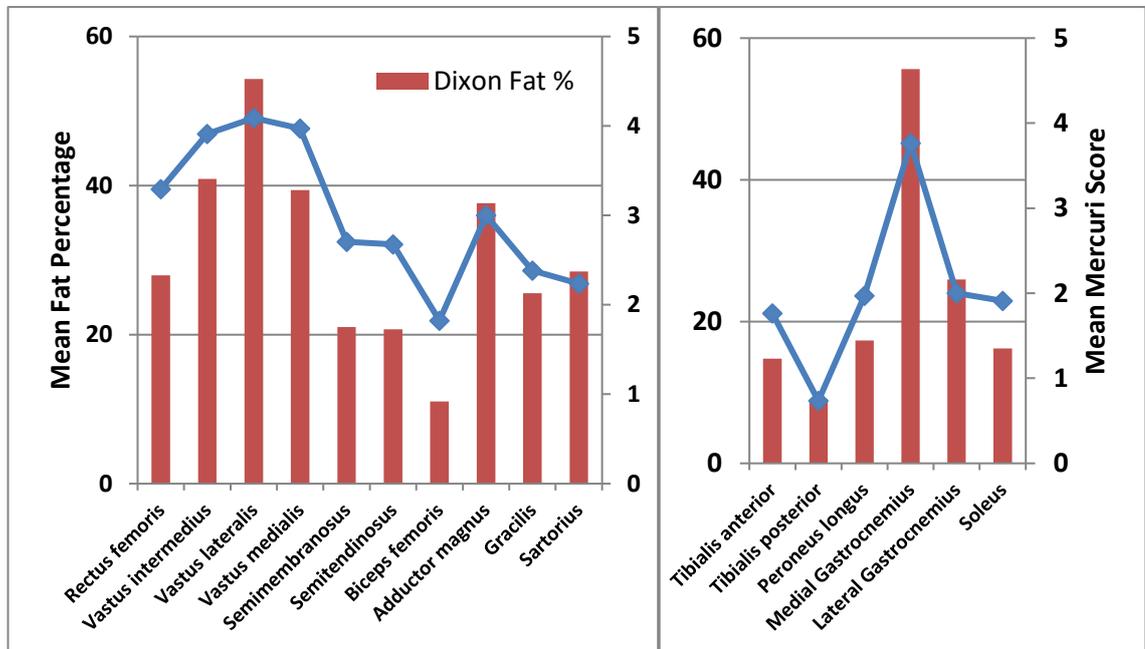
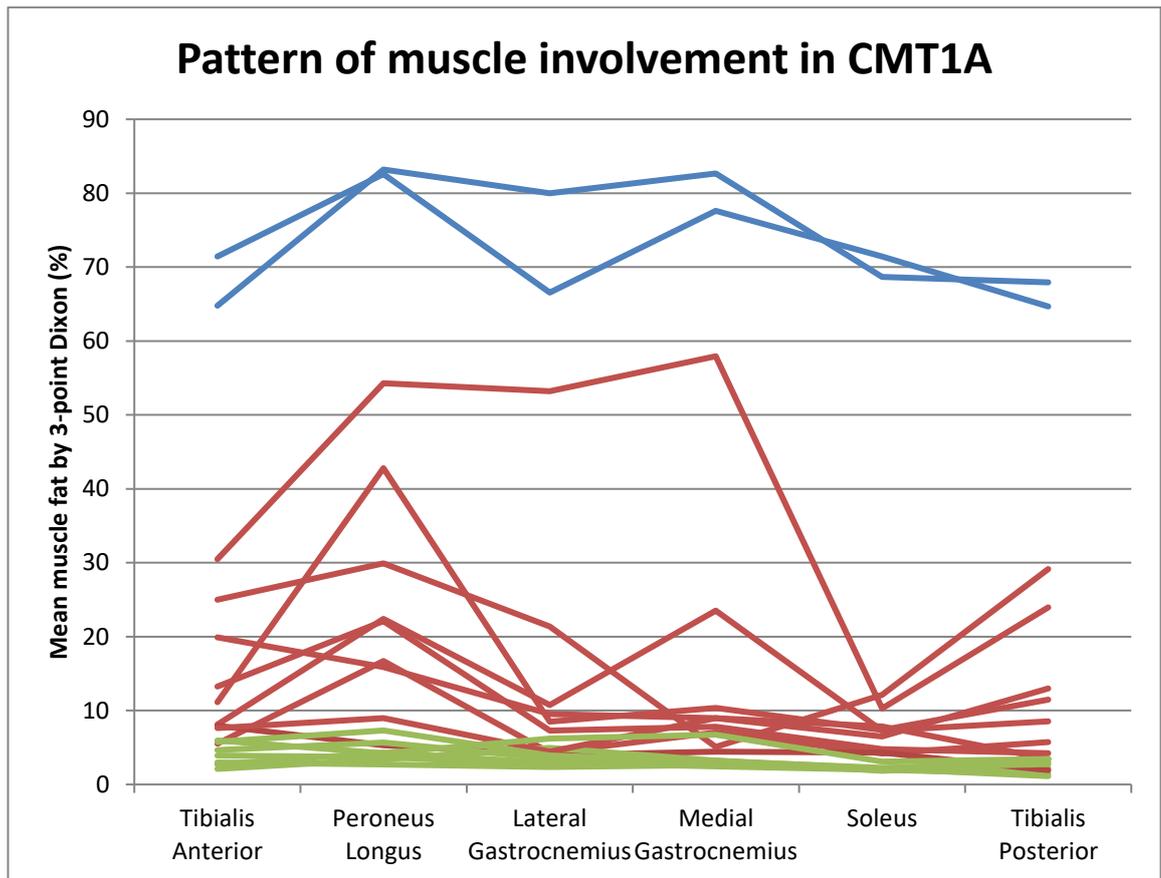


Figure 4-5: Distribution of intramuscular fat accumulation in calf muscles of CMT1A

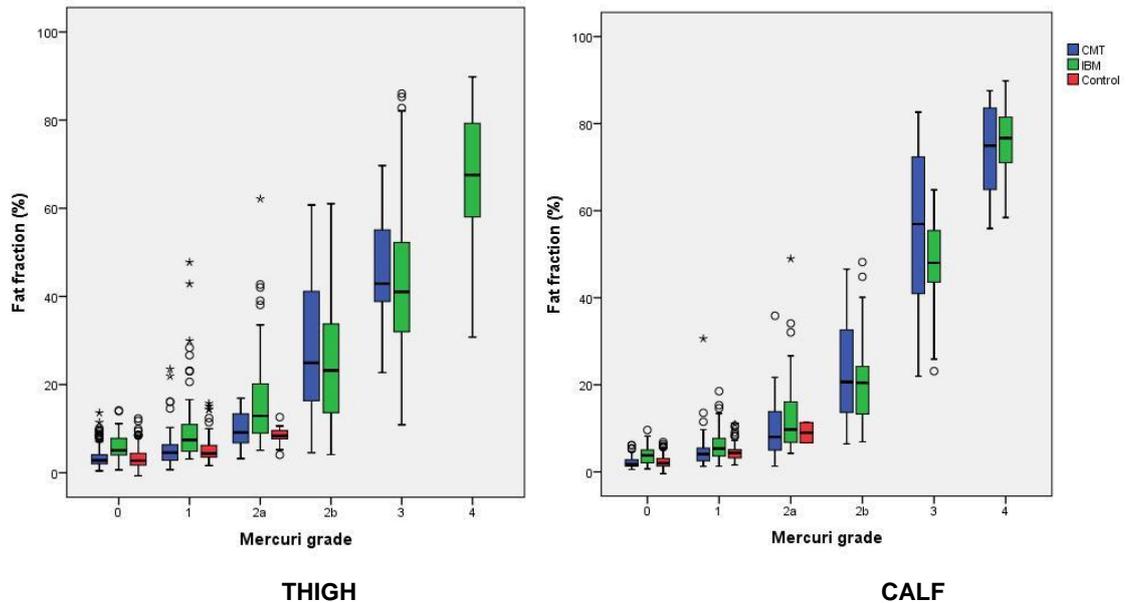


Each line represents a single patient. Right and left legs are averaged for each muscle. Lines are coloured into high (blue), intermediate (red) and low (green) levels of fat fraction as a visual aid.

#### 4.4.3.4 MRI correlations

There were strong correlations between mean Mercuri grades and mean slice fat fraction (Figure 4-6). There was however significant overlap in FF recorded for adjacent Mercuri grades. FF values for the same Mercuri grade were similar in all three subject groups.

Figure 4-6: Correlation between FF and Mercuri grade



*Box plot of fat fraction measured according to Mercuri grade assigned in thigh and calf muscles, grouped by diagnosis. Measured fat fraction increases with higher Mercuri grades and is consistent across thigh and calf and diagnoses. Bars indicate median, 25<sup>th</sup>, 50<sup>th</sup> and 75<sup>th</sup> centiles, lines: range, °: minor outlier, \*: major outlier*

There were strong correlations between the quantitative MRI parameters of individual muscles (Table 4-7). All correlations were greater than  $R = 0.89$ , and all highly significant ( $p < 0.0001$ ). The strongest correlations are between T2 and MTR at both thigh and calf level. T1 overall has the weakest correlations. In the thigh, at a nominal 0% fat fraction T1 is 1267ms, T2 is 42.6ms and MTR is 31.8 p.u. For every 1% increase in FF, T1 reduces by 12.5ms, MTR by 0.38 p.u. and T2 increases by 1.35ms. Inter-parameter relationships are similar at calf level. Whilst T2 and MTR will have additional influence from water distribution changes and the relationship is not strictly linear, these are reasonable estimates of the confounding effect of intramuscular fat accumulation on T2 and MTR measurements.

Table 4-7: Correlation of quantitative MRI parameters

Thigh			Dependent variable			
			FF	T1	T2	MTR
Independent variable	FF	R		0.91	0.92	0.93
		Constant		1267	42.6	31.8
		Slope		-12.5	1.35	-0.38
	T1	R	0.91		0.89	0.93
		Constant	86		166	-1.8
		Slope	-0.067		-0.096	0.026
	T2	R	0.92	0.89		0.95
		Constant	-26	1610		42.2
		Slope	0.633	-8.343		-0.253
	MTR	R	0.93	0.93	0.95	
		Constant	73	220	155	
		Slope	-2.266	33.169	-3.548	

Calf			Dependent variable			
			FF	T1	T2	MTR
Independent variable	FF	R		0.95	0.93	0.96
		Constant		1265	41.6	31.7
		Slope		-11.7	1.19	-0.35
	T1	R	0.95		0.92	0.95
		Constant	98		162	-4.2
		Slope	-0.076		-0.095	0.028
	T2	R	0.93	0.92		0.97
		Constant	-29	1620		43.2
		Slope	0.73	-8.93		-0.28
	MTR	R	0.96	0.95	0.97	
		Constant	84	240	149	
		Slope	-2.613	31.927	-3.376	

Linear regression of pair-wise comparison of quantitative MRI parameters on an individual muscle basis at thigh and calf level. For example, at thigh level the equation  $T2 = 42.6 + 0.92 \times \text{fat fraction}$  fits data best, with  $R = 0.92$ . All quantitative MRI parameters are highly significant ( $p < 0.001$  for all). Constant and slope are similar at thigh and calf level for equivalent correlations. FF: fat fraction; MTR: magnetisation transfer ratio; R: model fit parameter.

#### 4.4.4 Probing water distribution abnormalities

The STIR sequence allows the most straightforward assessment of abnormalities in muscle water distribution due to the suppression of fat signal within this sequence, but is a qualitative sequence. MTR and T2 mapping are quantitative sequences sensitive to changes in muscle water distribution, but the confounding effect of intramuscular fat accumulation (Table 4-7) needs to be taken into consideration.

#### 4.4.4.1 Qualitative STIR sequences

Within the thigh muscles, in controls and CMT1A patients marked STIR hyperintensity was not seen and mild hyperintensity was uncommon. Conversely in IBM patients over a third of muscles are mildly hyperintense with an additional 20% markedly hyperintense on STIR. Distribution of calf STIR grades in IBM were roughly similar to the thigh, whilst 6.3% of CMT1A calf muscles were graded as showing marked STIR hyperintensity, versus 0.5 % of control calf muscles (Table 4-8).

Table 4-8: Frequency of STIR grades in thigh and calf muscles

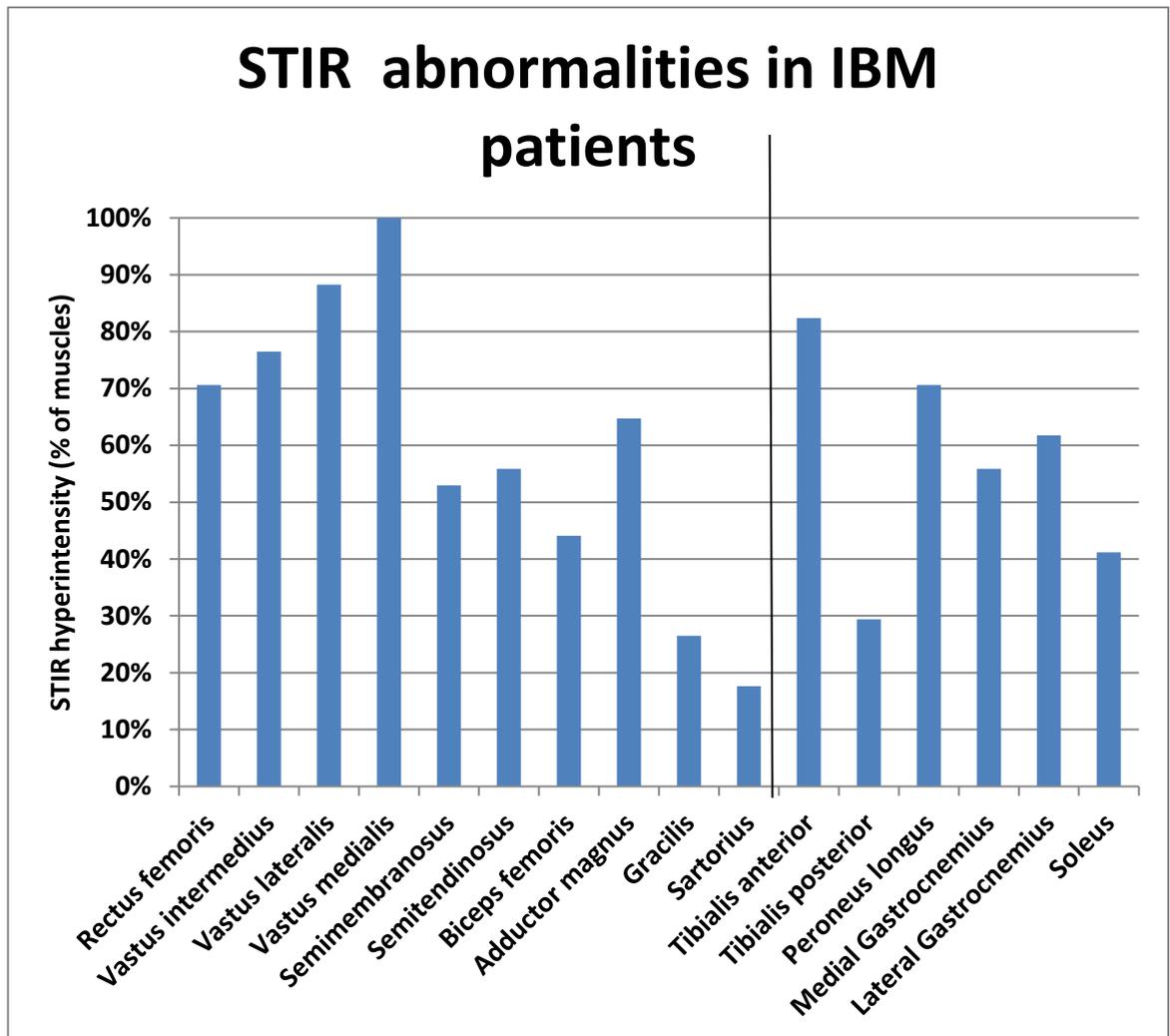
Thigh			STIR				Total
			None	Mild	Marked	Art.	
Diagnosis	CMT1A	Count	352	19	0	29	400
		% within Diagnosis	88.0%	4.8%	0.0%	7.3%	100%
	IBM	Count	176	141	80	3	400
		% within Diagnosis	44.0%	35.3%	20.0%	0.8%	100%
	Controls	Count	553	16	0	11	580
		% within Diagnosis	95.3%	2.8%	0.0%	1.9%	100%
Total		Count	1081	176	80	43	1380
		% within Diagnosis	78.3%	12.8%	5.8%	3.1%	100%

Calf			STIR				Total
			None	Mild	Marked	Art.	
Diagnosis	CMT1A	Count	162	60	15	3	240
		% within Diagnosis	67.5%	25.0%	6.3%	1.3%	100%
	IBM	Count	107	93	38	2	240
		% within Diagnosis	44.6%	38.8%	15.8%	0.8%	100%
	Controls	Count	304	49	2	17	372
		% within Diagnosis	81.7%	13.2%	0.5%	4.6%	100%
Total		Count	573	202	55	22	852
		% within Diagnosis	67.3%	23.7%	6.5%	2.6%	100%

The distribution of STIR abnormalities in IBM patients is shown in Figure 4-7. The pattern in the thigh is similar to the pattern of intramuscular fat accumulation except gracilis, which is relatively spared but moderately affected by fat. Of note, 100% of vastus medialis shows STIR hyperintensity.

Within the calf muscles, medial gastrocnemius which almost universally shows greatest fat accumulation has only the fourth highest frequency of STIR hyperintensity, with tibialis anterior the most commonly STIR affected muscle. Tibialis posterior least commonly shows STIR hyperintensity, congruent to its relative sparing from IFA.

Figure 4-7: Distribution of STIR abnormalities in IBM patients



#### 4.4.4.2 Early water abnormalities in muscles without significant intramuscular fat accumulation

As noted in section 4.4.3.4, the measures selected to quantify abnormal muscle water distribution in this study: T2 and MTR, are significantly influenced by intramuscular fat accumulation. One approach taken was to assess muscles without significant IFA, defined here as the 95<sup>th</sup> percentile of IFA levels of the muscles of control subjects. For the 29 control subjects, the FF 95<sup>th</sup> percentiles were 4.8% for thigh muscles and 4.7% for calf-level muscles, defining the upper thresholds for IFA in these healthy controls. Muscles with “normal” levels of intramuscular fat are also an important group to study as they represent an early phase in the disease process before irreversible muscle damage has occurred.

Within just this subset of muscles without significant intramuscular fat accumulation, significantly increased T2 and reduced MTR were seen in both thigh and calf muscles of both IBM and CMT patients compared with matched controls (Table 4-9). In the same subgroup, T2 and MTR measurement differed by qualitative STIR assessment:

increased STIR signal was associated with increased T2 and reduced MTR in both thigh and calf muscles in IBM patients and in calf muscles in CMT patients (Table 4-10). This association was also evident in the control group. Notably, T2 was significantly elevated in IBM patients at both thigh and calf level and in CMT1A patients at calf level, even when STIR assessment was normal. These T2 and MTR abnormalities prior to significant fat infiltration provide means to quantify early NMD pathology, with abnormalities quantifiable even in muscles without STIR hyperintensity.

*Table 4-9: Quantitative parameters in muscles without significant IFA*

Group	CMT	Control (CMT)	IBM	Control (IBM)
Thigh FF	1.5 ± 1.0*	1.4 ± 1.1	2.5 ± 1.2***	1.6 ± 1.1
Thigh T2	42.6 ± 4.7***	41.2 ± 4.5	48.7 ± 7.9***	42.5 ± 4.4
Thigh MTR	31.7 ± 1.9**	32.1 ± 1.5	29.6 ± 2.3***	31.7 ± 1.6
Calf FF	1.9 ± 1.1***	1.4 ± 1.0	2.2 ± 1.2***	1.6 ± 1.1
Calf T2	42.2 ± 4.6***	39.2 ± 3.0	45.5 ± 7.0***	40.9 ± 3.8
Calf MTR	31.4 ± 1.7***	32.3 ± 1.2	30.4 ± 2.3***	31.7 ± 1.3

Comparison of mean values of quantitative MRI parameters between patient and control groups. Significant differences (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ) are apparent even in these muscles without significant fat infiltration. CMT: Charcot-Marie-Tooth disease; IBM: inclusion body myositis; FF: fat fraction; MTR: magnetisation transfer ratio.

*Table 4-10: Quantitative parameters in muscles without significant IFA grouped by STIR*

Group	IBM		CMT		Control (All)	
	Normal	Increased	Normal	Increased	Normal	Increased
Thigh FF	2.3 ± 1.2	2.7 ± 1.2	1.5 ± 1.0	1.7 ± 1.1	1.4 ± 1.0	2.2 ± 1.4**
Thigh T2	46.6 ± 6.3	51.9 ± 9.0***	42.6 ± 4.6	42.7 ± 5.3	41.6 ± 4.3	44.3 ± 3.7*
Thigh MTR	30.1 ± 2.0	28.8 ± 2.5**	31.7 ± 1.9	31.6 ± 1.7	32 ± 1.6	31.1 ± 1.1*
Calf FF	2.0 ± 1.1	2.3 ± 1.2	1.7 ± 1.1	2.2 ± 1.2*	1.4 ± 1.0	1.6 ± 1.0
Calf T2	42.6 ± 3.5	48.2 ± 8.3***	41.4 ± 3.8	44.5 ± 5.7***	39.7 ± 3.6	41.8 ± 3.9***
Calf MTR	31.7 ± 1.3	29.2 ± 2.5***	31.7 ± 1.3	30.6 ± 2.4***	32.2 ± 1.3	31.6 ± 1.3***

Comparison of mean values of quantitative MRI parameters of muscles without significant fat infiltration according to qualitative findings on STIR imaging. Significant (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ) increases in T2 and decreases in MTR are seen in muscles with qualitative STIR hyperintensity in IBM, CMT and control groups.

However, even in this subgroup of muscles without significant intramuscular fat accumulation, patients had slightly higher FF than their respective controls (Table 4-9). To account for this, regression analysis was performed including muscle FF as a covariate. Regression analyses were performed separately for the IBM and CMT groups, combined with their respective controls, including only data from individual patient and control muscles with FF below the “normal” FF thresholds. On this analysis, T2 and MTR remained significantly dependent upon FF (Table 4-11);

however the dependencies for control versus disease status were also significant in each case, with increased T2 (thigh:  $+4.0 \pm 0.5$ ms [co-eff $\pm$ SE], calf:  $+3.5 \pm 0.6$ ms) and reduced MTR (thigh:  $-1.5 \pm 0.2$ p.u., calf:  $-1.1 \pm 0.2$ p.u.) independent of FF in this regression model, in IBM versus matched controls. Smaller but significant FF-adjusted dependencies were also observed in CMT1A versus matched-controls, especially at calf-level (T2 thigh:  $+1.0 \pm 0.3$ ms; T2 calf:  $+2.0 \pm 0.3$ ms; MTR thigh  $-0.3 \pm 0.1$ p.u.; MTR calf  $-0.7 \pm 0.1$ p.u.).

Table 4-11: Linear regression T2 and MTR in muscles fat fraction in the healthy control range

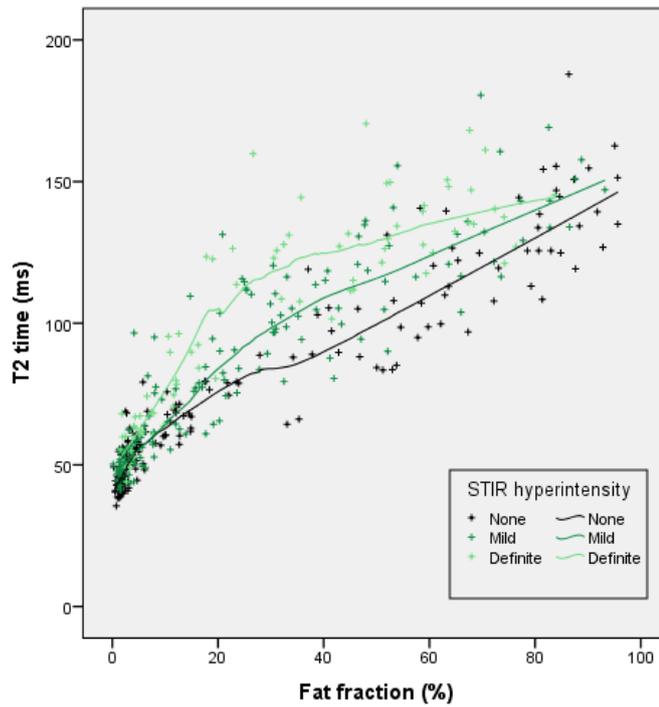
<b>A</b>	<b>Thigh</b>			<b>Calf</b>		
<b>T2</b>	R=0.62, p<0.001			R=0.52, p<0.001		
	<i>Co-eff</i>	<i>Std Error</i>	<i>p</i>	<i>Co-eff</i>	<i>Std Error</i>	<i>p</i>
Constant	38.4	0.43	<0.0001	38.0	0.53	<0.0001
IBM	4.0	0.54	<0.0001	3.5	0.57	<0.0001
FF	2.52	0.21	<0.0001	1.78	0.25	<0.0001
<b>MTR</b>	R=0.61, p<0.001			R=0.39, p<0.001		
	<i>Co-eff</i>	<i>Std Error</i>	<i>p</i>	<i>Co-eff</i>	<i>Std Error</i>	<i>p</i>
Constant	32.9	0.16	<0.0001	32.3	0.20	<0.0001
IBM	-1.5	0.19	<0.0001	-1.1	0.21	<0.0001
FF	-0.74	0.07	<0.0001	-0.33	0.09	0.0007
<b>B</b>	<b>Thigh</b>			<b>Calf</b>		
<b>T2</b>	R=0.49, p<0.001			R=0.59, p<0.001		
	<i>Co-eff</i>	<i>Std Error</i>	<i>p</i>	<i>Co-eff</i>	<i>Std Error</i>	<i>p</i>
Constant	38.4	0.29	<0.0001	36.6	0.31	<0.0001
CMT1A	1.0	0.31	0.0008	2.0	0.33	<0.0001
FF	2.07	0.15	<0.0001	1.87	0.17	<0.0001
<b>MTR</b>	R=0.40, p<0.001			R=0.38, p<0.001		
	<i>Co-eff</i>	<i>Std Error</i>	<i>p</i>	<i>Co-eff</i>	<i>Std Error</i>	<i>p</i>
Constant	33.0	0.12	<0.0001	32.8	0.14	<0.0001
CMT1A	-0.3	0.13	0.04	-0.7	0.15	<0.0001
FF	-0.65	0.06	<0.0001	-0.40	0.08	<0.0001

Muscles with FF < 4.8% thigh; FF < 4.7% calf in IBM (A) and CMT1A (B) patients and matched controls. Determinants of T2 and MTR in muscles with normal FF. FF remains strongly correlated to both T2 and MTR, however subject group (patient = 1, control =0) also has a significant effect. R: Overall model correlation coefficient; CMT1A: Charcot-Marie-Tooth disease type 1A; IBM: inclusion body myositis; FF: fat fraction; MTR: magnetisation transfer ratio; Co-eff: partial regression coefficient; p: significance level.

#### 4.4.4.3 Assessment of T2 values relative to FF levels

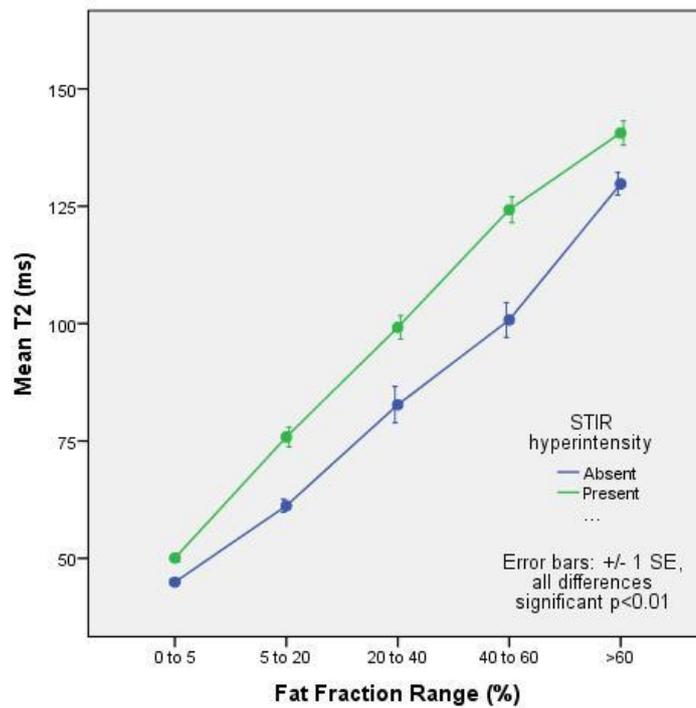
The relationship between FF and T2 for higher levels of FF was also assessed. Within thighs of IBM patients, T2 was higher for equivalent FF in muscles with STIR hyperintensity (Figure 4-8). This is demonstrated on a scatter plot and also by grouping FF within the ranges: <5%, 5-20%; 20-40%, 40-60% and >60% (Figure 4-9).

Figure 4-8: T2 vs Fat Fraction in thigh muscles of IBM patients grouped by qualitative STIR rating



Loess best fit lines for different STIR ratings are shown.

Figure 4-9: Thigh FF range vs T2 in IBM patient grouped by STIR positivity



#### 4.4.5 Clinical data

Baseline clinical measures are summarised in Table 4-12. CMT1A patients had a median age of onset of 6 years, ranging from 1 year to a 27 year old asymptomatic patient (diagnosed on family testing). The CMT patients covered a broad range of severity, with CMT-examination score ranging from 0-18/28 (mean 8.0, standard deviation 5.1). In IBM subjects, the median IBM-FRS was 28.5, range 18-35, so were overall of mild to moderate severity.

*Table 4-12: Clinical data summary*

Measure	CMT	IBM
<b>Age of onset</b>	6.0 ± 4.4	59.0 ± 8.5
<b>Disease duration</b>	35.8 ± 17.5	7.7 ± 3.1
<b>CMTSS (0-12)</b>	3.1 ± 2.0	NA
<b>CMTES (0-28)</b>	8.0 ± 5.1	NA
<b>MRC LL (0-110)</b>	95.4 ± 15.4	93.4 ± 15.7
<b>MRC UL (0-120)</b>	112.4 ± 7.9	95.2 ± 12.8
<b>MRC Total (0-230)</b>	207.8 ± 21.8	188.6 ± 27.1
<b>SF36 (0-100%)</b>	73.9 ± 15.2	61.5 ± 15.3
<b>SF36-PF (0-100%)</b>	65.3 ± 23.2	39.5 ± 19.1
<b>IBMFRS-LL (0-12)</b>	NA	5.9 ± 2.4
<b>IBMFRS (0-40)</b>	NA	27.6 ± 5.4

*MRC: Medical Research Council bedside strength assessment; SF-36: Short Form Health Survey; PF: physical functioning domain; CMTES: CMT Examination Score; IBM-FRS: IBM Functional Rating Score; NA: Not applicable.*

##### 4.4.5.1 Myometry

In both patient groups the baseline myometric muscle strength was significantly reduced in all muscle groups compared with their matched controls (Table 4-13 and Table 4-14, all  $p < 0.01$ ).

*Table 4-13: Baseline myometry statistics by group*

Measure	CMT	Control (CMT)	p	IBM	Control (IBM)	p
<b>Knee extension</b>	93.6 ± 44.1	134.9 ± 43.5	0.005	26.8 ± 26.0	119.9 ± 43.0	<0.0001
<b>Knee flexion</b>	47.2 ± 20.1	66.1 ± 20.4	0.006	35.3 ± 19.9	60.6 ± 20.1	<0.0001
<b>Ankle plantarflexion</b>	26.0 ± 14.3	62.0 ± 19.1	<0.0001	29.6 ± 15.7	51.6 ± 18.4	<0.0001
<b>Ankle dorsiflexion</b>	10.8 ± 7.5	30.0 ± 9.8	<0.0001	13.2 ± 10.8	28 ± 10.6	<0.0001

*Values are mean ± standard deviation in Nm of all isometric and isokinetic measurements combined for each movement.*

As expected, in CMT patients, weakness was greater for ankle movements (calf muscles). In CMT overall ankle dorsiflexion was, relative to controls, weaker than ankle plantarflexion as per the classical clinical description (Reilly et al., 2011); however this difference was only observed in isometric strength assessment (at 10

degrees plantarflexion) whereas with isokinetic strength measurement both ankle plantarflexion and ankle dorsiflexion were similar, roughly half of control values when assessed. Furthermore, the angle of peak isokinetic strength in CMT1A patients was at a significantly greater degree of plantarflexion than in the matched controls (30 degrees versus 10%), suggesting that the reduced isometric strength observed for ankle dorsiflexion may be partly due to the angle chosen for assessment combined with reduced range of ankle dorsiflexion. A similar pattern was seen for ankle eversion which was weaker than ankle inversion, especially with isometric assessment. The observed knee weakness in CMT1A patients is somewhat unexpected, given the general lack of fatty infiltration observed in the thigh muscles, and in 18/20 CMT1A patients, knee movements were graded full strength on bedside examination (MRC grade 5). This does concur with a small reduction in thigh muscle size observed in CMT1A patients (Table 4-2 and Table 4-4), and may therefore be related to deconditioning due to a reduction in general physical activity compared with the healthy controls in the study

In IBM patients weakness was greater for knee movements (thigh muscles), especially knee extension, as expected from the clinical phenotype (Hilton-Jones et al., 2010). All ankle movements were also weak, with marginally greater weakness of ankle eversion. Unlike the CMT1A patients, isokinetic and isometric reductions in strength versus matched controls were of similar magnitude in IBM patients.

Table 4-14: Baseline isokinetic and isometric myometry by group

Movement	Type	Side/angle	CMT	Control (CMT)	IBM	Control (IBM)
Knee Extension	Isometric	Right at 45°	111.1 ± 48.9 (24-209)	155.6 ± 47.3 (73-259)	35.4 ± 38.4 (0-146)	138.6 ± 60 (18-259)
		Left at 45°	109.7 ± 47.6 (38-217)	152.2 ± 45.5 (73-231)	33.5 ± 38.3 (0-145)	138.9 ± 48.7 (62-218)
		Right at 90°	112.2 ± 58.4 (45-290)	159.3 ± 62.6 (84-292)	29.2 ± 23.9 (2-91)	136.6 ± 55.4 (18-252)
		Left at 90°	106.6 ± 60.8 (43-277)	142 ± 50.5 (72-247)	28.3 ± 26.5 (0-94)	128.5 ± 43.3 (72-216)
	Isokinetic	Right at 60°/s	87 ± 45 (12-184)	133.5 ± 49.1 (49-239)	25 ± 22 (0-87)	116.8 ± 45.4 (18-190)
		Left at 60°/s	87.4 ± 42.1 (23-165)	127.6 ± 40 (52-202)	24.2 ± 24.4 (0-94)	115.7 ± 38 (52-174)
		Right at 120°/s	70 ± 32.5 (27-138)	106.9 ± 40.8 (41-206)	20.2 ± 18.3 (0-71)	93 ± 38.5 (13-170)
		Left at 120°/s	68.6 ± 35.3 (26-138)	102.5 ± 35.2 (45-165)	19 ± 19.7 (0-80)	91.4 ± 32.8 (45-145)
Knee Flexion	Isometric	Right at 45°	65.8 ± 26.9 (33-115)	86.6 ± 28.9 (53-141)	49.6 ± 27.4 (2-110)	80 ± 29.1 (33-140)
		Left at 45°	66.3 ± 27.5 (33-113)	81.8 ± 22.9 (35-132)	48.8 ± 25.5 (2-106)	74.9 ± 27.7 (35-132)
		Right at 90°	47.9 ± 20.4 (14-92)	64.7 ± 27.1 (31-127)	28.9 ± 18.1 (0-68)	54.7 ± 18.9 (26-106)
		Left at 90°	45.4 ± 20.9 (15-83)	57.8 ± 24.3 (28-110)	31.3 ± 20.9 (0-77)	49.3 ± 18.3 (24-95)
	Isokinetic	Right at 60°/s	46.6 ± 18.8 (9-81)	70 ± 21.7 (28-123)	34.9 ± 21.4 (0-81)	64.5 ± 21 (27-99)
		Left at 60°/s	45.1 ± 20 (8-85)	62.9 ± 17.3 (20-85)	37 ± 21.5 (0-80)	62.4 ± 22.7 (20-104)
		Right at 120°/s	35.1 ± 16.8 (14-72)	55.5 ± 19.7 (22-102)	26.1 ± 17.7 (0-72)	50.8 ± 17.8 (22-87)
		Left at 120°/s	34.5 ± 18.6 (12-75)	50 ± 14.9 (20-77)	25.9 ± 17 (0-57)	48.5 ± 18.2 (20-84)
Ankle Plantar-flexion	Isometric	Right at 10°	33.7 ± 19.6 (1-64)	66.6 ± 20.8 (35-106)	34.4 ± 18.7 (3-64)	56.2 ± 19.1 (28-92)
		Left at 10°	32.7 ± 19.8 (0-64)	65.1 ± 19.5 (33-108)	33.6 ± 16.7 (4-60)	57.2 ± 23.7 (33-110)
	Isokinetic	Right at 60°/s	18.7 ± 12.8 (5-58)	57.7 ± 21.6 (24-107)	24.1 ± 14.3 (0-54)	46.1 ± 17 (22-75)
		Left at 60°/s	19 ± 12.4 (4-53)	58.6 ± 21.7 (22-115)	29.2 ± 17.2 (7-71)	46.9 ± 20.5 (8-79)
Ankle Dorsi-flexion	Isometric	Right	9.7 ± 8.1 (0-31)	35.7 ± 13.7 (16-60)	15 ± 12.9 (0-38)	33 ± 14.9 (7-60)
		Left at 10°	10 ± 7.8 (0-27)	34.5 ± 13.1 (12-58)	15.4 ± 12.5 (0-38)	32.6 ± 14.2 (8-58)
	Isokinetic	Right at 60°/s	11.9 ± 8.4 (0-30)	25.4 ± 8.5 (12-42)	11.8 ± 12.5 (0-50)	23.3 ± 9.2 (9-42)
		Left at 60°/s	11.9 ± 11 (0-43)	24.5 ± 8.7 (7-37)	11.8 ± 8.3 (0-26)	23.2 ± 9 (7-35)
Ankle Inversion	Isometric	Right at 0°	15.6 ± 10 (0-37)	20.4 ± 7.1 (9-33)	11.9 ± 7.3 (0-27)	18.1 ± 6.4 (9-33)
		Left at 0°	14.8 ± 10.4 (0-45)	20.1 ± 7.1 (7-30)	11.5 ± 6.7 (0.15-26)	17.7 ± 6.7 (7-30)
	Isokinetic	Right at 60°/s	16 ± 9.9 (0-38)	25.4 ± 7.9 (11-39)	14.9 ± 9.6 (3-34)	22.7 ± 7.8 (9-34)
		Left at 60°/s	14.3 ± 9.5 (3-35)	25.6 ± 8.9 (11-43)	17.1 ± 10.2 (3-35)	21.8 ± 8.6 (11-39)
Ankle Eversion	Isometric	Right at 0°	7.1 ± 4.1 (1-14)	22 ± 7.6 (9-37)	9.6 ± 5.9 (0-22)	19 ± 7.6 (9-34)
		Left at 0°	7.7 ± 4.8 (0-18)	21.4 ± 7.1 (11-34)	10.2 ± 6.5 (0-20)	17.5 ± 6.9 (8-31)
	Isokinetic	Right at 60°/s	8.4 ± 4 (1-15)	17.3 ± 5.6 (9-28)	9.2 ± 5 (0-20)	15.1 ± 5.6 (8-23)
		Left at 60°/s	7.8 ± 3.1 (3-14)	17.4 ± 5.8 (9-30)	9.3 ± 5.8 (0-24)	16.2 ± 7.1 (8-33)

Data are mean ± standard deviation (range). Isometric values are the peak torque at the fixed angle listed whilst isokinetic values are the peak torque at the fixed speed noted. Both CMT1A patients and IBM patients have significantly ( $p < 0.01$  for all) reduced strength than their matched control groups for all measurements.

#### 4.4.6 MRI - clinical correlations

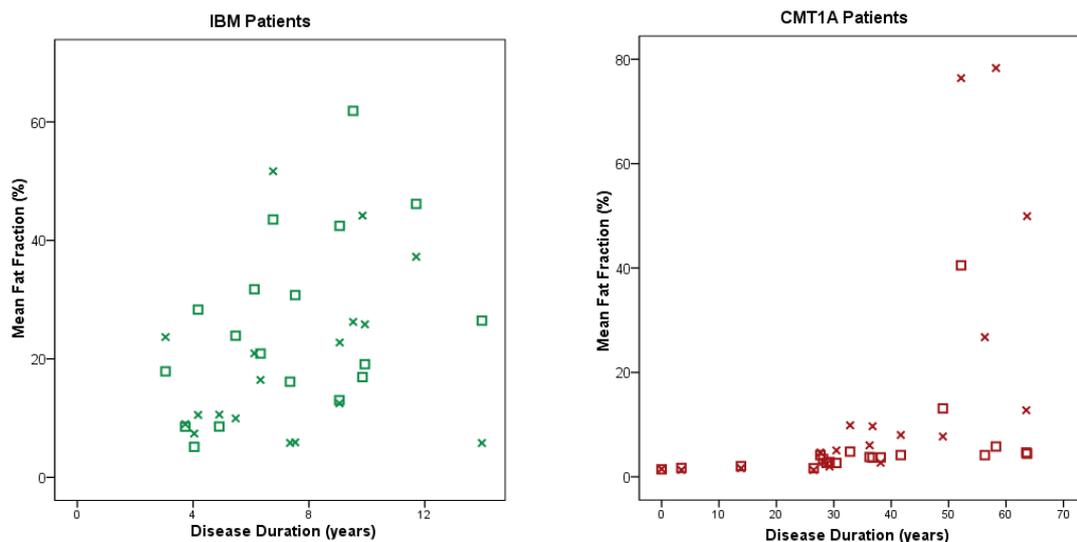
There were significant correlations in both CMT1A and IBM patient groups between overall clinical measures and summary MRI FF measurements, and between specific myometric strength measurements and quantitative MRI measurements in relevant functional groups.

#### 4.4.6.1 Correlations between overall MRI and clinical measures

In IBM patients, all-muscle thigh-level FF correlated with disease duration ( $\rho=0.50$ ,  $p=0.03$ , Figure 4-10), IBM-FRS ( $\rho=-0.53$ ,  $p=0.02$ ), IBM-FRS lower limb components ( $\rho=-0.64$ ,  $p=0.002$ , Figure 4-11), the total MRC score lower limb ( $\rho=-0.60$ ,  $p=0.005$ ), and the physical function domain of the SF36 ( $\rho=-0.60$ ,  $p=0.007$ ). There were not significant correlations with overall SF36 ( $\rho=-0.11$ ,  $p=0.66$ ) or age ( $\rho=0.14$ ,  $p=0.55$ ).

In CMT, all-muscle calf-level FF correlated with disease duration ( $\rho=0.89$ ,  $p<0.0001$ , Figure 4-10), age ( $\rho=0.84$ ,  $p<0.0001$ , Figure 4-12), CMTES ( $\rho=0.63$ ,  $p=0.003$ , Figure 4-13), lower limb motor component of the CMTNS ( $\rho=0.77$ ,  $p<0.0001$ ) and reduced total MRC score lower limb ( $\rho=-0.76$ ,  $p<0.0001$ ). The correlation between all-muscle calf FF and total SF36 score was not significant ( $\rho=-0.34$ ,  $p=0.18$ ), however the correlation with the SF-36 physical function domain score was significant ( $\rho=-0.63$ ,  $p=0.007$ ).

Figure 4-10: Correlation of fat fraction with disease duration in IBM and CMT1A patients



Scatter plot of disease duration versus thigh-level ( $\square$ ) and calf-level ( $\times$ ) fat fraction in IBM and CMT1A patients. In IBM patients positive correlation between disease duration and mean fat fraction is seen at thigh level ( $\rho=0.50$ ,  $p=0.03$ ) but not calf level ( $\rho=0.27$ ,  $p=0.28$ ). In CMT1A patients, strong positive correlations at both levels are seen (thigh:  $\rho=0.81$ ,  $p<0.0001$ ; calf:  $\rho=0.89$ ,  $p<0.0001$ ). The relationship between disease duration and fat fraction appears non-linear in the CMT1A group.

Figure 4-11: Correlation between IBM-FRS (lower limb) and thigh FF in IBM patients

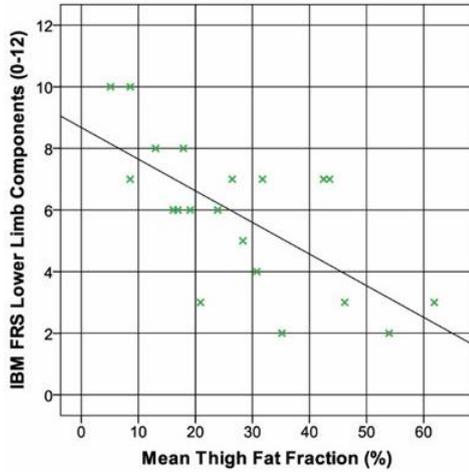


Figure 4-12: Relationship between age and calf FF in CMT1A patients

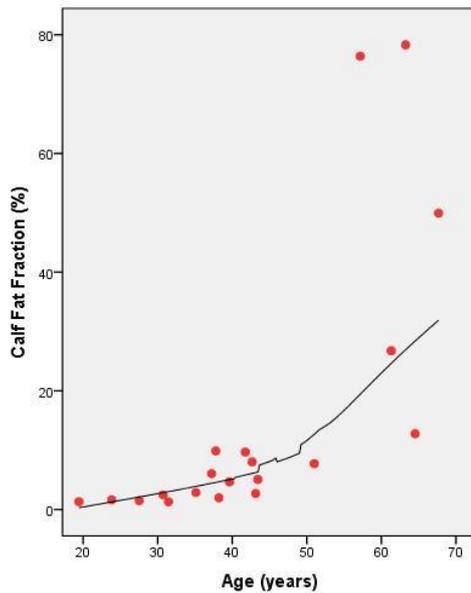
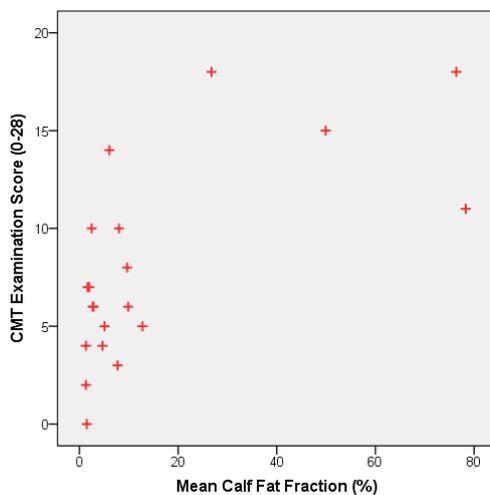


Figure 4-13: Correlation of calf fat fraction with severity score (CMTES) in CMT1A patients



There is significant correlation between calf muscle fat fraction and CMTES in CMT1A patients ( $\rho=0.63$ ,  $p=0.003$ ). The relationship appears non-linear.

Figure 4-14: Correlation matrix of MRI FF and IBM-FRS domains in IBM patients

		Correlations															
		FF_LowerLimb	FF_Thigh	Swallow	handwriting	cutting food/using utensils	fine motor tasks	dressing	hygiene	turning in bed/adjusting covers	sit to stand	walking	Climbing stairs	total	LL total	Thigh total	
Spearman's rho	FF_LowerLimb	Correlation Coefficient	1.000	.872*	-.184	.044	-.084	-.058	.049	-.058	-.213	-.523	-.611*	-.654	-.366	-.437	-.598*
		Sig. (2-tailed)	.	.000	.437	.855	.724	.809	.837	.809	.366	.018	.004	.002	.112	.054	.005
		N	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
FF_Thigh		Correlation Coefficient	.872*	1.000	-.265	.026	-.017	-.041	-.065	-.224	-.299	-.679**	-.592*	-.636*	-.481*	-.567*	-.694*
		Sig. (2-tailed)	.000	.	.259	.913	.943	.864	.785	.342	.200	.001	.006	.003	.032	.009	.001
		N	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
Swallow		Correlation Coefficient	-.184	-.265	1.000	.138	.163	.257	.179	-.330	.159	.375	.538	.474	.506	.465	.462
		Sig. (2-tailed)	.437	.259	.	.563	.491	.275	.449	.155	.503	.103	.014	.035	.023	.039	.028
		N	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
handwriting		Correlation Coefficient	.044	.026	.138	1.000	.702*	.398	.183	.475*	.493	.092	.364	.279	.570*	.429	.284
		Sig. (2-tailed)	.855	.913	.563	.	.001	.082	.440	.034	.027	.699	.114	.234	.009	.059	.224
		N	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
cutting food/using utensils		Correlation Coefficient	-.084	-.017	.163	.702*	1.000	.676*	.243	.357	.467*	-.035	.300	.143	.515	.287	.093
		Sig. (2-tailed)	.724	.943	.491	.001	.	.001	.301	.122	.038	.883	.198	.548	.020	.220	.696
		N	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
fine motor tasks		Correlation Coefficient	-.058	-.041	.257	.398	.676*	1.000	.669*	.297	.490	.332	.404	.316	.695*	.521*	.365
		Sig. (2-tailed)	.809	.864	.275	.082	.001	.	.001	.203	.028	.152	.078	.175	.001	.018	.114
		N	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
dressing		Correlation Coefficient	.049	-.065	.179	.183	.243	.669*	1.000	.550	.474	.474	.164	.297	.813*	.635*	.340
		Sig. (2-tailed)	.837	.785	.449	.440	.301	.001	.	.012	.035	.035	.489	.203	.004	.003	.142
		N	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
hygiene		Correlation Coefficient	-.058	-.224	.330	.475*	.357	.297	.550	1.000	.552*	.468	.247	.465	.692*	.739*	.400
		Sig. (2-tailed)	.809	.342	.155	.034	.122	.203	.012	.	.007	.030	.294	.039	.001	.000	.081
		N	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
turning in bed/adjusting covers		Correlation Coefficient	-.213	-.299	.159	.493	.467*	.490*	.474*	.582*	1.000	.553	.479	.332	.714*	.655*	.482
		Sig. (2-tailed)	.366	.200	.503	.027	.038	.028	.035	.007	.	.011	.033	.153	.000	.002	.031
		N	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
sit to stand		Correlation Coefficient	-.523*	-.679**	.375	.092	-.035	.332	.474*	.486*	.553*	1.000	.627*	.659*	.740*	.833*	.835*
		Sig. (2-tailed)	.018	.001	.103	.699	.883	.152	.035	.030	.011	.	.003	.002	.000	.000	.000
		N	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
walking		Correlation Coefficient	-.611*	-.592*	.538	.364	.300	.404	.164	.247	.479	.627*	1.000	.746*	.734*	.724*	.882*
		Sig. (2-tailed)	.004	.006	.014	.114	.198	.078	.489	.294	.033	.003	.	.000	.000	.000	.000
		N	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
Climbing stairs		Correlation Coefficient	-.654*	-.636*	.474	.279	.143	.316	.297	.466*	.332	.659*	.746*	1.000	.742*	.830*	.875*
		Sig. (2-tailed)	.002	.003	.035	.234	.548	.175	.203	.039	.153	.002	.000	.	.000	.000	.000
		N	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
total		Correlation Coefficient	-.366	-.481*	.506	.570*	.515*	.695*	.613*	.692*	.714*	.740*	.734*	.742*	1.000	.939*	.796*
		Sig. (2-tailed)	.112	.032	.023	.009	.020	.001	.004	.001	.000	.000	.000	.000	.	.000	.000
		N	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
LL total		Correlation Coefficient	-.437	-.567*	.465	.429	.287	.521*	.635*	.739*	.655*	.833*	.724*	.830*	.939*	1.000	.879*
		Sig. (2-tailed)	.054	.009	.039	.059	.220	.018	.003	.000	.002	.000	.000	.000	.000	.	.000
		N	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
Thigh total		Correlation Coefficient	-.598*	-.694*	.492	.284	.093	.365	.340	.400	.482	.835*	.882*	.875*	.796*	.879*	1.000
		Sig. (2-tailed)	.005	.001	.028	.224	.696	.114	.142	.081	.031	.000	.000	.000	.000	.000	.
		N	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20

\*\* . Correlation is significant at the 0.01 level (2-tailed).  
 \* . Correlation is significant at the 0.05 level (2-tailed).

It was informative to assess the cross-correlation matrix in the IBM patients between lower limb FF, thigh FF and the individual domains which make up the IBM-FRS (Figure 4-14). Although thigh fat fraction did correlate with the overall IBM-FRS ( $\rho=0.48$ ,  $p=0.03$ ), it did not significantly correlate with the swallowing domain or any of the domains which predominantly assess upper limb function. Indeed, even between the IBM-FRS components, the upper limb components generally correlated with each other, and the lower limb components correlated with each other, but there was no correlation between upper and lower limb components.

#### 4.4.6.2 Correlation between MRI measures and myometry

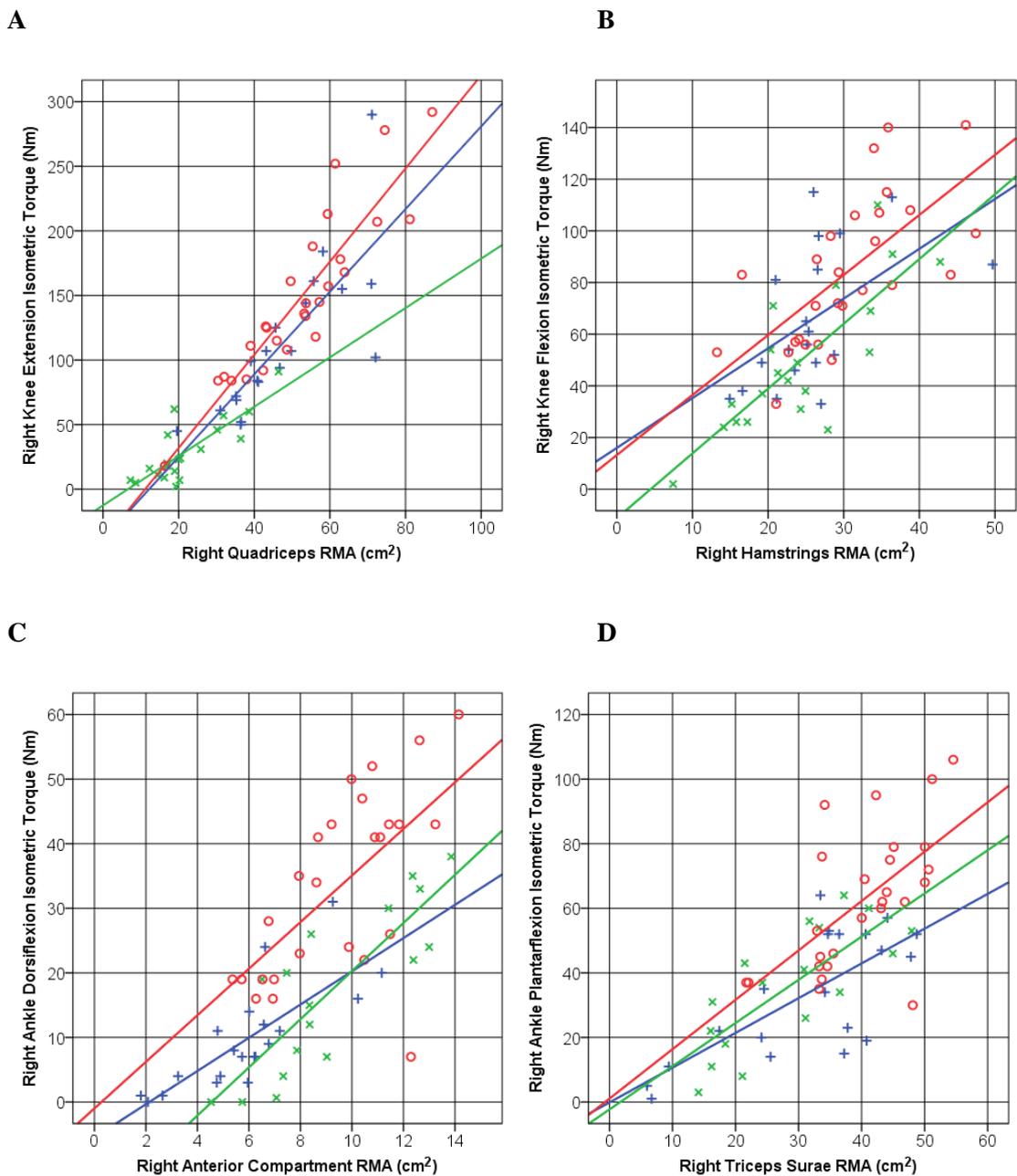
In all subgroups for all movements assessed, muscle strength showed strong positive correlation with total cross-sectional area of the corresponding muscle group (Table 4-15). In regions where patient fat fraction was elevated (IBM thigh and calf level, CMT calf level), muscle fat fraction (whole ROI) showed strong negative correlation with muscle strength (Table 4-15). The combined parameter “remaining muscle area” (an overall measure of fatty atrophy) had the strongest correlation with muscle strength (Figure 4-15, Table 4-15).

Table 4-15: Correlation of muscle strength with MRI parameters in each subject group

		IBM			CMT			Control		
		Area	FF	RMA	Area	FF	RMA	Area	FF	RMA
Right Quadriceps	Pearson ICC	0.665**	-0.551*	0.809**	0.780**	-0.323	0.791**	0.903**	-0.407*	0.907**
	p	0.001	0.012	0.000	0.000	0.165	0.000	0.000	0.035	0.000
Left Quadriceps	Pearson ICC	0.727**	-0.778**	0.917**	0.848**	-0.286	0.860**	0.900**	-0.251	0.905**
	p	0.000	0.000	0.000	0.000	0.222	0.000	0.000	0.207	0.000
Right Hamstrings	Pearson ICC	0.702**	-0.561*	0.800**	0.469*	-0.252	0.539*	0.647**	-0.264	0.670**
	p	0.001	0.010	0.000	0.043	0.298	0.017	0.000	0.183	0.000
Left Hamstrings	Pearson ICC	0.706**	-0.579**	0.790**	0.587**	-0.295	0.664**	0.635**	-0.247	0.660**
	p	0.001	0.007	0.000	0.008	0.220	0.002	0.000	0.215	0.000
Right Anterior Compartment	Pearson ICC	0.780**	-0.509*	0.833**	0.607**	-0.544*	0.775**	0.600**	-0.195	0.622**
	p	0.000	0.037	0.000	0.004	0.013	0.000	0.001	0.339	0.001
Left Anterior Compartment	Pearson ICC	0.675**	-0.538*	0.796**	0.606**	-0.634**	0.759**	0.672**	-0.376	0.677**
	p	0.002	0.021	0.000	0.005	0.004	0.000	0.000	0.058	0.000
Right Triceps Surae	Pearson ICC	0.476*	-0.626**	0.769**	0.531*	-0.609**	0.719**	0.589**	-0.235	0.613**
	p	0.039	0.007	0.000	0.016	0.004	0.000	0.001	0.248	0.001
Left Triceps Surae	Pearson ICC	0.452	-0.556*	0.681**	0.741**	-0.681**	0.834**	0.493**	-0.010	0.540**
	p	0.052	0.017	0.002	0.000	0.001	0.000	0.008	0.963	0.004

IBM: inclusion body myositis; CMT: Charcot-Marie-Tooth disease; FF: fat fraction; RMA: remaining muscle area (see methods for definition); Pearson ICC: Pearson inter-class correlation coefficient; \* significant correlation  $p < 0.05$ ; \*\* significant correlation  $p < 0.01$ .

Figure 4-15: Correlation of MRI measured remaining muscle area with isometric muscle strength



MRI measured RMA of lower limb muscle groups show strong correlation with corresponding strength in IBM (x), CMT1A (+) and controls (o). A: Right quadriceps. IBM:  $\rho=0.81$ ,  $p<0.0001$ ; CMT:  $\rho=0.79$ ,  $p<0.0001$ ; controls:  $\rho=0.91$ ,  $p<0.0001$ . B: Right hamstrings. IBM:  $\rho=0.80$ ,  $p<0.0001$ ; CMT:  $\rho=0.54$ ,  $p=0.02$ ; controls:  $\rho=0.67$ ,  $p<0.0001$ . C: Right anterior compartment. IBM:  $\rho=0.83$ ,  $p<0.0001$ ; CMT:  $\rho=0.78$ ,  $p<0.0001$ ; controls:  $\rho=0.62$ ,  $p=0.0007$ . D: Right triceps surae. IBM:  $\rho=0.77$ ,  $p<0.0001$ ; CMT:  $\rho=0.72$ ,  $p<0.0001$ ; controls:  $\rho=0.61$ ,  $p=0.0009$ .

### 4.4.6.3 Assessment of muscle quality

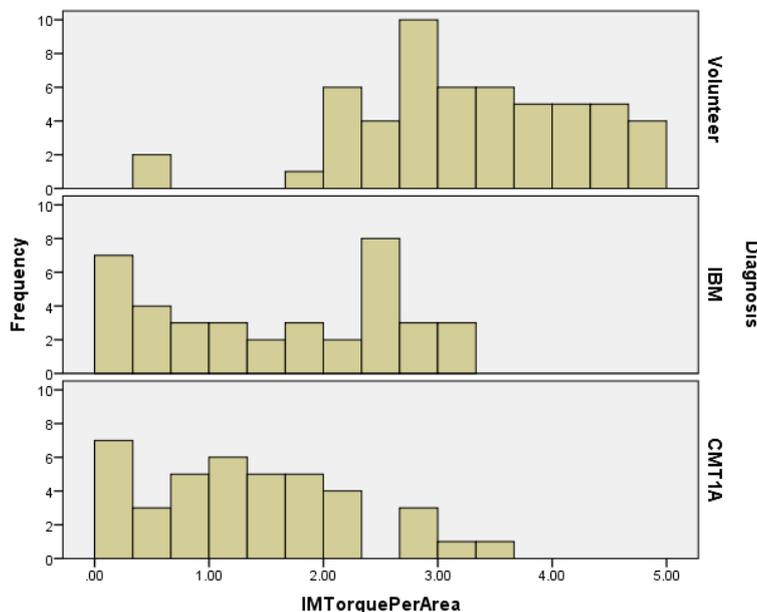
To account for the confounding effect of differences in muscle size, some papers define a metric for “muscle quality” as being the maximum torque generated per cross-sectional area:

$$\text{Specific Strength} = \text{Strength} / \text{Muscle Cross-sectional Area}$$

*Equation 4-1: Muscle quality formula*

This type of assessment has been used in measuring the effect of ageing (Delmonico et al., 2009; Jubrias et al., 1997), longitudinally in diabetic neuropathy (Andreassen et al., 2009) and also in myotonic dystrophy (Hiba et al., 2012). Reduced muscle quality, beyond even the effect of intramuscular fat accumulation can be seen in Figure 4-15, where the patient best fit lines for the remaining muscle area to strength correlation is lower than that in healthy controls, demonstrating that patient muscles are weaker even when the fatty atrophy is taken into account. The distribution of muscle quality (isometric torque per unit cross-sectional area) is shown in Figure 4-16 with clearly reduced muscle quality in both IBM and CMT1A patients compared with controls. There is one outlier in the control group.

*Figure 4-16: Muscle quality of tibialis anterior in the three subject groups*



*Distribution of tibialis anterior muscle quality, defined as maximum isometric ankle dorsiflexion torque (Nm) per cm<sup>2</sup> of total muscle area across the three subject groups.*

A potential advantage of this assessment is that it allows analysis of correlation with any of the quantitative parameters, by using the value derived from the small ROI for T1/T2/MTR and the cross-sectional muscle area from the full ROI. Thus one can assess if MTR or fat fraction estimates “muscle quality” better or if the addition of T2 values adds to estimation of “muscle quality” from fat fraction alone.

Correlations between isometric and isokinetic torque per unit area and MRI parameters for tibialis anterior muscle are shown in Table 4-16. There is a negative correlation with age and BMI (decreasing muscle quality as age or BMI increase) and a positive correlation with height. All quantitative parameters showed strong, highly significant correlations with muscle quality. Whole ROI FF had stronger correlations than small ROI FF, but the best performing quantitative parameter was MTR followed by T2 time – the measures which are sensitive to both chronic and acute pathology. Considering just the IBM patients, who had greatest degree of acute changes quantifiable, MTR and age provided a more accurate model of muscle quality than FF and age (Figure 4-17).

Table 4-16: Correlates of tibialis anterior muscle quality

Factor	Isometric torque per area		Isokinetic torque per area	
	Pearson ICC	p-value	Pearson ICC	p-value
Age (y)	-0.29	0.001	-0.39	<0.001
Height (cm)	0.26	0.003	0.06	0.47
Weight (kg)	0.01	0.91	-0.13	0.13
BMI (kg/m <sup>2</sup> )	-0.18	0.04	-0.23	0.009
FF-small (%)	-0.52	<0.001	-0.50	<0.001
T1 time (ms)	0.53	<0.001	0.53	<0.001
T2 time (ms)	-0.62	<0.001	-0.59	<0.001
MTR (p.u.)	0.66	<0.001	0.63	<0.001
FF-whole (%)	-0.57	<0.001	-0.54	<0.001

Parametric correlation of tibialis anterior specific force (Nm/cm<sup>2</sup>) with quantitative MRI and demographic parameters in all subjects

Figure 4-17: Regression analyses of tibialis anterior muscle quality for IBM patients

MTR and age	R=0.78, p=<0.001		
	Co-eff	Std Error	p
Constant	0.43	1.17	<0.001
Age	-0.030	0.010	<0.001
MTR	0.148	0.030	<0.001

FF and age	R=0.68, p<0.001		
	Co-eff	Std Error	p
Constant	5.33	0.61	<0.001
Age	-0.042	0.010	<0.001
FF	-0.051	0.017	<0.001

The dependent variable is maximum isometric torque per cross-sectional area in Nm/cm<sup>2</sup>, a measure of “muscle quality”. Age is significantly related. The addition of MTR (top) measurement results in a more accurate model than the addition of FF (bottom).

## **4.5 Discussion**

In the cross-sectional phase of the study, fat infiltration strongly correlated with strength, function and overall severity, demonstrating high criterion validity for these measurements. T2 and MTR were abnormal in muscles without significant fat infiltration quantifying early and potentially reversible pathological changes. In addition to specific insights into CMT1A and IBM, the published literature demonstrates that these methods will have wide applicability across a broad spectrum of neuromuscular diseases.

### **4.5.1 Validity of MRI measures of fatty atrophy**

Demonstration of validity is the central aim of this cross-sectional study phase. These data provide direct support for both criterion and construct validity of quantitative MRI measures in these representative neuromuscular diseases. In addition, the data suggest which measures will have best content validity for these diseases.

#### **4.5.1.1 Construct validity**

Construct validity is the ability of a test to measure the underlying concept of interest to the researcher. This was assessed in this study by both the known groups method and through demonstration of convergent validity.

##### **4.5.1.1.1 Known groups method**

The known groups method assesses whether the measurement differs in groups known to differ in the construct of interest. In this study we can consider the quantitative MRI measurements between the CMT1A group and their matched controls, between the IBM group and their matched controls, and also directly between the CMT and IBM patient groups.

The MRI measures in the CMT1A group are all significantly different from the matched controls at calf level but not thigh level (Table 4-2). All measures are similarly affected due to the well-known strong interdependencies of these measures, demonstrated again in this study (Table 4-7), an example of convergent validity. There was also measurable muscle atrophy at the calf level (Table 4-2), consistent with clinical observations in this disease (Reilly et al., 2011). A small effect on T2 and MTR independent of and predating intramuscular fat accumulation was seen at calf level (Table 4-9, Table 4-11), demonstrating construct validity for these methods in detecting early abnormalities relating to muscle denervation.

In the IBM group, greater changes in quantitative parameters were seen at thigh than calf level (Table 4-2), again consistent with the observed clinical phenotype (Hilton-Jones et al., 2010). The pattern of involvement within the thigh with predominant affliction of quadriceps (Table 4-6) is also one of the key clinical features of IBM, and

thus provides construct validity. Significantly abnormal T2 and MTR in muscles without significant fat infiltration is consistent with inflammatory changes seen on muscle biopsy. Again the ability to demonstrate the clinically expected pattern of involvement provides construct validity for these measurements. These findings are consistent with previous studies which have applied fat quantification methods to neuromuscular diseases (4.2.2).

This cross-sectional study is unique in including a large number of patients with two completely dissimilar neuromuscular diseases. Most previous quantitative studies have included either a single disease, or occasionally a large number of diseases each with only a few subjects. This study allows assessment of construct validity through direct comparison of results between the two patient groups. Again the measurements are as would be expected from these diseases: a greater degree of fatty atrophy in thigh muscles in IBM than in CMT1A, similar levels of IFA in the calf muscles, and overall greater T2/MTR abnormalities independent of fat infiltration in the IBM group.

#### **4.5.1.1.2 Convergent validity**

Convergent validity is demonstrated when results on the test being administered are similar to those thought to measure the same underlying construct. In this study the correlation observed between established qualitative MRI methods and novel quantitative MRI methods is a demonstration of convergent validity.

Consistent with other studies (Willis et al., 2014; Wokke et al., 2013) there are clear relationships between qualitative Mercuri grading of IFA and quantitative FF in the same muscle (Figure 4-6) and the relationship between STIR abnormality and measure T2 and MTR (Figure 4-8, Figure 4-9, Table 4-10).

Although the qualitative and quantitative measures show a clear relationship, the advantages of the quantitative assessment are clear. Consistent with previous studies, the qualitative gradings overestimate the fat content compared with that qualitatively measured. For example, grade Mercuri grade 2b is defined as 30-60% fat replacement, but the median FF of muscles scored 2b was 24 (IQR 17-37). Similarly, the majority of muscles graded 3 scored less than the 60% FF designated as the lower boundary of this category. Significant overlap is seen between adjacent Mercuri grade categories, and wide ranges of FF within one Mercuri grade. Both of these emphasise that qualitative grading is more useful for diagnostic than outcome measure purposes, which will be directly assessed using longitudinal data in chapter 5. Qualitative data however does provide sufficient information for patient stratification. This is already used for muscle biopsy site selection where normal muscles or severely affected

muscles are avoided, and may prove useful at a screening phase in a clinical trial should a homogenous patient group be desired.

#### **4.5.1.2 Criterion validity**

Criterion validity is assessment of the novel measurement against a gold standard, and is perhaps the most important demonstration of validity in a medical context. In particular outcome measure validity should be demonstrated through correlation to relevant patient function. This is vital for both drug licencing agencies and the patient.

We demonstrated strong clinical-MRI correlations for both overall subject and individual muscle measures. In both CMT1A and IBM there are strong correlations between overall MRI measures and quality of life indices, functional or composite scales (IBMFRS and CMTES) and bedside strength examination, consistent with previous research (Hiba et al., 2012; Huang et al., 1994; Kim et al., 2010a; Willis et al., 2013). As might be expected, correlations were stronger with more closely related functional measures – for example mean fat fraction correlated with the physical function domain of the quality of life score, but not the total score; for CMT1A patients within the CMTES, correlation was strongest with the lower limb motor function score, and for IBM patients correlation was strongest with the lower limb motor relevant aspects of the IBM-FRS (sit-to stand, walking and stair climb). Indeed, there was no significant correlation between lower limb FF and function in the swallowing or upper limb domains – which may have been expected on an overall disease severity basis. This suggests that upper limb and lower limb function progress independently, and may therefore have different pathogenic mechanisms. Conversely in CMT1A, correlations are seen between FF and all aspects of the CMTES, including domains such as sensory impairment not directly related to what MRI has measured, suggesting that pathogenesis of this disease is more homogenous, with more uniform patterns of disease progression.

At the individual muscle level, the intuitively expected correlation between CSA and strength in healthy controls was demonstrated here as previously (Maughan et al., 1983). In patients there was an additional negative correlation between strength and FF, demonstrated previously in myotonic dystrophy for ankle dorsiflexion (Hiba et al., 2012) and here for both ankle and knee movements in both diseases. Muscle CSA and level of fat infiltration combined as the “remaining muscle area” metric correlated most strongly with strength. Thus MRI provides indices of chronic muscle pathology which are highly correlated to muscle strength, but independent of subject effort or operator involvement which lead to the poor test-retest and inter-observer reliability (Solari et al., 2008) of direct muscle strength measures. MRI therefore provides a valid, reliable surrogate measure of muscle strength.

### **4.5.1.3 Face validity**

Both reduction in muscle size and replacement of muscle with fat have excellent face validity as being measures of disease severity. MRI-measured muscle fat has previously been shown to correlate strongly with biopsy-measured muscle fat with a correlation coefficient of 0.995 (Gaeta et al., 2011). As the structure of skeletal muscle is relatively straightforward, reduction of contractile cross-sectional area (remaining muscle area) through fat replacement or muscle atrophy is a very plausible measure of disease severity. Conversely T2, MTR and T1 measurements have less face validity, and their validity requires greater justification through the other means mentioned.

### **4.5.1.4 Content validity**

Content validity is the degree to which the measurement includes all aspects relevant to the concept being measured. This needs to be considered for each disease separately.

In CMT1A the distribution of weakness is length dependent, and hence predominantly lower limb, although upper limb weakness does occur at later stages of the disease (Reilly et al., 2011). On the one hand, only muscle abnormalities are detectable with the MRI protocol used and the protocol will not directly measure sensory deficit, which is an important feature of CMT1A. On the other hand, calf muscle fat fraction correlated with all components of the CMTES – including the sensory symptoms, sensory signs and upper limb weakness domains. Although correlation was highest for the lower limb motor domain, correlation with all domains suggests some overall uniformity in disease progression across the domains, and an accurate measure of one domain may then be considered an appropriate marker of overall disease severity.

The validity of a pure motor measurement in CMT1A will also be dependent on the intervention being assessed. Lower limb MRI would not have content validity for a drug which benefited only sensory nerves, or a physiotherapy intervention targeting only upper limb muscles. If a lower limb muscle assessment is considered valid, a second issue is, does a single slice at thigh level and a single slice at calf level have content validity – or should whole muscle measurements be performed. A particular consideration is the ceiling effect, where a number of patients with CMT1A (green lines in Figure 4-5) have similar intramuscular fat at the calf level to healthy controls, whereby MRI quantification of intramuscular fat does not adequately define disease severity in these patients, although the findings of T2 and MTR abnormalities before IFA may provide some coverage for this group of patients. A potential solution would be to measure more distally in the calf, or measure the foot musculature, both of which are planned in future projects. Ultimately, poor content validity will have a negative impact on outcome measure responsiveness, and is considered further in Chapter 5.

Patients with IBM classically have a distinctive pattern of involvement: quadriceps weakness, poor finger grip and difficulty swallowing. A lower limb MRI protocol is only able to measure the first of these. The IBM-FRS includes questions which address all these domains. Whilst the lower limb MRI indices correlate with total IBM-FRS as well as the lower limb domains, they did not correlate with the upper limb or swallowing domains, and further the upper limb domains did not show significant correlation with the lower limb domains (Figure 4-14). An intervention which improved upper limb function independent of lower limb function would not be measurable with the protocol in this study. A protocol including both upper limb and lower limb muscles would have greater content validity. Other authors (Cantwell et al., 2005; Cox et al., 2011a; Phillips et al., 2001; Sekul et al., 1997) have demonstrated intramuscular fat accumulation on qualitative T1-weighted sequences of forearm muscles, most marked in flexor digitorum profundus. Together with colleagues at the Institute of Child Health, we have quantified forearm muscle fat fraction in Duchenne muscular dystrophy (Ricotti et al., 2014), so it would be straightforward to improve content validity by including coverage of forearm muscle groups if desired.

Beyond the simple metrics such as muscle fat fraction or muscle size, this thesis has examined combined metrics which might be considered to have superior content validity. Remaining muscle area is a metric which is influenced both by muscle size and intramuscular fat accumulation, and demonstrated strongest correlation with muscle strength. Whilst T2 and MTR could be criticised as being influenced by both acute water changes and chronic intramuscular fat accumulation, they have higher content validity since they encompass both processes, and notably had stronger correlations to muscle quality (specific muscle force) than fat fraction alone (Table 4-16).

#### **4.5.1.5 Validity conclusions**

Quantitative MRI measures of fatty atrophy have high validity in these exemplar neuromuscular diseases. Most importantly they demonstrate criterion validity through excellent correlation with patient strength and function in both CMT1A and IBM. They also show strong construct validity both through known groups method assessment and convergent validity. They also have clear face validity and content validity, with the only potential concern in this regard that for IBM, only the lower limbs were imaged, though the same fat quantification methods can and have been applied to the upper limbs.

#### **4.5.2 Measures of abnormal muscle water distribution**

Even after adjusting for residual FF, T2 was increased and MTR reduced in muscles without significant IFA in both patient groups compared to controls, suggesting

sensitivity to early and potentially reversible oedematous changes. Since the T2 of fat markedly exceeds that of muscle-tissue water, and *vice versa* for MTR, non-fat-suppressed T2 and MTR are influenced by both FF, exemplified by the strong inter-MRI parameter correlations, and potentially independent early tissue water distribution changes. For this reason T2 obtained in this manner has been referred to as *total T2* (Carlier, 2014). Our data from muscles with FF below the 95<sup>th</sup> percentile of the control range suggest that adjusting for T2 and MTR dependence upon residual FF reveals significant differences between patient and control groups independent of FF. These differences, greater for the IBM group but also significant in CMT1A, may reflect early changes in muscle water distribution occurring prior to significant IFA. These changes may be reversible with effective therapy (Jurkat-Rott et al., 2009), and muscle T2 and MTR may thus provide useful biomarkers in clinical trials focussed upon early or active disease. There are now methods to separate the fat and water components of T2 at the acquisition stage (Janiczek et al., 2011). These sequences have longer acquisition times but may be included when abnormal water measurement is a focus of the study.

Whilst T2 and MTR have lower face validity than fat fraction assessment, they may have superior content validity as they are measures of both acute and chronic muscle pathology. Relative T2 and MTR measures in IBM and CMT1A versus controls matches the pattern expected by disease phenotype, demonstrated by construct validity through the known groups methods. Correlations between STIR hyperintensity and increased T2, and between STIR hyperintensity and reduced MTR (Table 4-10, Figure 4-8, Figure 4-9) are a demonstration of the convergent validity of these methods. Finally, MTR appears a better measure than FF of muscle quality, i.e. predictor of muscle strength, a demonstration of criterion validity. This work therefore, even with the limitations of the methods used, provides a number of lines of evidence for the validity of T2 and MTR in neuromuscular diseases.

### **4.5.3 Insights into CMT1A**

Although utilised here as an exemplar neuromuscular disease, this study provides specific insights into CMT1A, the commonest inherited neuromuscular disease. As expected from the clinical phenotype, the disease distribution observed was very much distal predominant, with only 3/20 patients having clear abnormalities in quantitative MRI parameters at the thigh level (full ROI thigh FF >5%). All these patients had severe changes of all muscles at calf level. Whilst distal predominance is generally true for most CMT subtypes, this absolute gradient is not always the case, for example qualitative MRI in CMT2F with HSPB1 mutations (Gaeta et al., 2012).

At the other end of the spectrum seven patients had normal fat fraction in all muscles at the calf level analysed. This is consistent with previous qualitative MRI reports in

CMT1A where in the mildest patients, abnormalities were only evident in the foot muscles (Berciano et al., 2008; Chung et al., 2008; Gallardo et al., 2006). As the protocol used was insufficient to show abnormalities, a protocol including quantitative foot imaging would be needed to fully characterise this group of patients. The relationship between age and calf intramuscular fat accumulation (Figure 4-12) is of particular interest in this regard. Albeit from a small sample size, CMT1A patients less than 35 had no significant fat accumulation at the calf level imaged, with intermediate levels of IFA in patients 35-55 and high levels in patients >55. This is counter to the notion put forward by Verhamme and colleagues (Verhamme et al., 2009) that the majority of disease progression in CMT1A occurs in childhood and that changes in function during adulthood are due to normal ageing processes with reduced reserve. It also suggests that MRI studies in children should focus on foot musculature, unless inclusion criteria target the more severe end of the spectrum.

The distribution of fat infiltration within muscles of the mid-calf when involvement is intermediate is worthy of consideration (see red lines in Figure 4-5). Similar to previous qualitative reports (Chung et al., 2008; Gallardo et al., 2006), generally peroneus longus shows maximum intramuscular fat accumulation, often followed by tibialis anterior. However, in some patients, medial gastrocnemius is the most affected muscle, which is clinically undetectable as soleus is universally one of the least affected muscles in patients with intermediate levels of calf involvement. As all fat quantification was performed at the same anatomical level, the differential effect is something other than location of the muscle in a proximal-distal axis, but may perhaps relate to different localisation of motor endplates on calf muscles and therefore different axon length. A small number of cases had significant asymmetry, sometimes in association with a clear cause such as ankle arthrodesis, and presumably in the other cases due to unidentified environmental factors such as soft tissue injuries. The greater affliction of ankle everters than ankle inverters has been proposed as important in the pathogenesis of pes cavus and other foot deformities. A study specifically looking at the relationship between MRI pattern of muscle involvement and foot deformity would certainly be of interest.

The main purpose of inclusion of myometry in the cross-sectional study was to assess validity for the MRI measures. However, isokinetic myometry has not previously been reported in CMT1A and provides some valuable insights. First, despite the length-dependent phenotype, there was a consistent reduction in knee extension and flexion strength for both isometric and isokinetic measurements in CMT1A patients compared with age matched controls (see Table 4-13). Averaged across all measurement methods, CMT1A patients were 31% weaker in knee extension and 29% weaker in

knee flexion than matched controls. It is notable that although this difference was highly significant ( $p < 0.01$ ), due to the wide range of normal strength in healthy controls, this weakness was not apparent on an individual patient basis, especially on manual muscle testing where almost all knee movements were graded as normal (grade 5). This discrepancy is partly explained by a 10% reduced cross-sectional area of thigh muscles in CMT1A patients versus matched controls, a difference which may be due to differences in activity levels between patients and controls and which may be reversed by an appropriate exercise and strengthening programme. Part of the reduced strength in the overall group relates to the increase fat fraction in thigh muscles of a small subset of the cohort (3/20) patients, but there remains a deficit in “muscle quality” (maximum force per unit cross-sectional area, seen where the correlation line between strength and remaining muscle area in patients runs below that of healthy controls).

A reduction in measured muscle quality is even more marked at calf level (see Figure 4-15, Figure 4-16). There are a number of plausible explanations for this finding, including muscle pathology distal to the level analysed, or mechanical considerations due to reduced range of motion of the ankle joint for example, which may also be amenable to some improvement with intervention. These mechanical effects are also the likely explanation for the relative isokinetic preservation (50% of control strength) of ankle eversion versus isometric at a neutral position (33% of control strength). This also highlights some of the difficulties in measuring calf muscle strength where there is a significant influence of ankle and foot deformity.

Finally, whilst all the discussion above relates to parameters of chronic denervation (fatty atrophy), which is not unexpected given the slowly progressive nature of this disease, there was also evidence of acute denervation on the water-sensitive sequences. On STIR, at the calf level there was a greater proportion of muscles graded mildly or markedly hyperintense than in controls (Table 4-8, mild 25.0% vs 13.2%, marked 6.3% vs 0.5%). Such occasional changes have been reported in CMT1A previously, in 4/11 patients in one series (Gallardo et al., 2006). In this thesis we have been able to quantify this using T2 and MTR sequences, with consideration of the confounding effect of fat infiltration. Specifically, through considering only calf muscles with less than 4.7% FF, muscles of CMT1A patients had significantly longer T2 and smaller MTR than matched controls (Table 4-9), which was still significant in regression analysis to account for any residual effect of small differences in FF (Table 4-11).

The magnitude of the differences observed was as expected smaller than in IBM patients (T2: +2.0ms CMT, +3.5ms IBM; MTR -0.7pu CMT, -1.1pu IBM) but the observation was highly significant ( $p < 0.0001$ ). Within this same subgroup of muscles

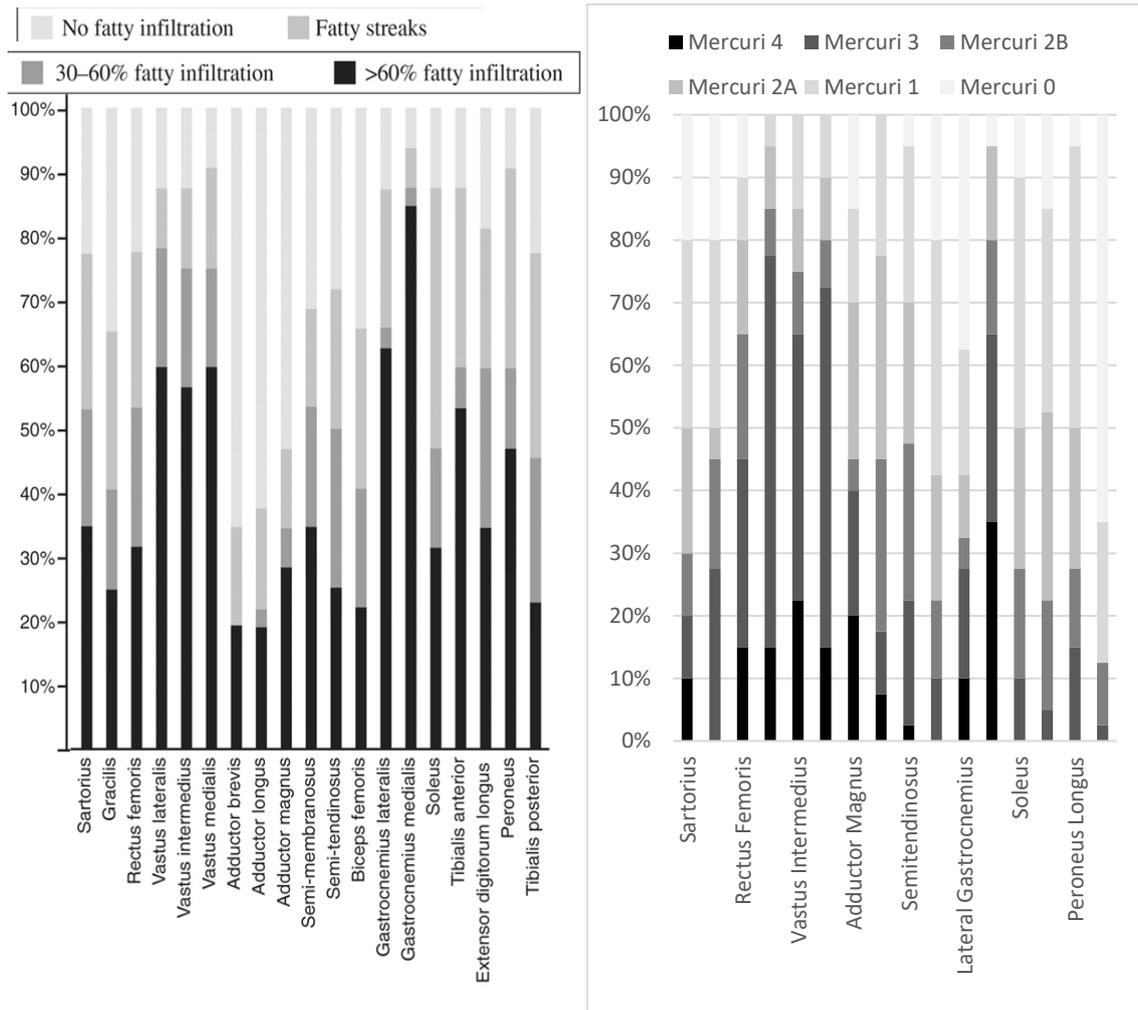
without significant fat infiltration, STIR positive muscle had longer T2 and smaller MTR in CMT1A patients, IBM patients and controls. However, a significant difference was observed even in the muscle without STIR hyperintensity (Table 4-10) demonstrating that these quantitative techniques can demonstrate abnormalities even in normal appearing muscle, which may be helpful in judging early treatment effects which reverse acute denervation. There are now advanced methods of T2 (Janiczek et al., 2011) and MTR (Sinclair et al., 2010) estimation which can separate the effects of fat and water at the point of acquisition and are certainly worthy of further investigation in CMT1A, even in this most “chronic” of diseases.

#### **4.5.4 Insights into IBM**

Inclusion body myositis was chosen as a neuromuscular disease with little commonality with CMT1A, but many interesting insights may still be gained from this detailed cross-sectional data set. Unlike in CMT1A, all MRI scans in IBM patients were qualitatively and quantitatively abnormal, with greater frequency and degree of acute disease, as might be expected with muscle inflammation being one of the hallmarks of disease pathology, and a greater rate of progression than in CMT1A.

In terms of the pattern of fat infiltration, this quantitative study showed findings similar to previous qualitative reports. Figure 4-18 shows the pattern from the previous largest series of 32 patients (Cox et al., 2011a), with data from this thesis presented in the same way. Within the thigh muscles, the overall pattern is very similar to previous reports: maximal affliction of quadriceps with relative sparing of rectus femoris within this group. The slight differences are that biceps femoris was relatively spared within hamstrings of patients reported in this thesis but not in the previous series, and adductor magnus, which is reported to be spared in the previous publication but shows a very wide range of severity in the patients reported in this thesis. The pattern within the calf muscles is identical to that previously reported, with medial gastrocnemius most affected, tibialis posterior least affected, and the remainder of muscles variably affected. It should be noted that as medical gastrocnemius works as part of the larger functional group triceps surae to plantarflex the ankle, this MRI disease pattern is not evident in the clinical phenotype, whereas predilection for quadriceps involvement is recognised as a key clinical feature. This common pattern raises the question as to whether MRI may have utility in IBM diagnosis, a question addressed in another project performed in parallel to this thesis.

Figure 4-18: Distribution of IFA in IBM in the literature (left) and this thesis (right)



This thesis also reports detailed isometric and isokinetic lower limb myometry of patients with IBM for the first time. At thigh level, the knee extension is weakest, approximately 20-25% of the strength of matched controls, which is consistent across different joint angles (isometric) and speed of movement (isokinetic). Knee flexion is also significantly weak, broadly similar with isometric (63% of control value) and isokinetic assessment (52% of control value). All ankle movements were also significantly weaker than controls with a similar degree of weakness (roughly 50%) of ankle plantarflexion, dorsiflexion and inversion, with some sparing of ankle inversion (70% of control value), consistent with the relative sparing of tibialis posterior on MRI. Degree of ankle weakness was more variable than at the knee, with some patients having normal strength and others close to no strength in all ankle movements. The pattern of weakness has important consequences on gait and function, and this relationship would be of interest in future studies.

Muscle inflammation is a key aspect of muscle histology in IBM, and although its relevance and usefulness as a therapeutic target is debated, it may be quantified on MRI. Overall 55% of muscles from IBM patients were at least mildly hyperintense on

STIR imaging at both thigh and calf level, which far exceeds that previously reported with a median of two muscles per patient (Cox et al., 2011b) using a whole body protocol. This difference might be explained by reduced sensitivity of their sequence due to the wider coverage, or excess sensitivity of our sequence, given the 13.7% rate of STIR positivity in calf muscles of our controls. This STIR positivity however was rare within the thigh muscles of controls, where patient abnormalities were as frequent, and is demonstrated to represent a quantifiable difference in muscle tissue MRI parameters even in healthy controls (see 3.4.7), so is not artefactual. In IBM patients the distribution of STIR abnormalities mirrored that of T1w abnormalities in the thigh, but in the calf tibialis anterior showed STIR abnormalities most commonly, with medial gastrocnemius only fourth most commonly. The explanation for this may be that the muscle tissue within medial gastrocnemius is end-stage, and therefore is no longer showing STIR positivity, which may have predominantly occurred when the patient was pre-symptomatic. The temporal relationships between STIR positivity and intramuscular fat accumulation are explored using longitudinal observation in chapter 5.

This thesis quantifies abnormal muscle water through T2 and MTR measurements, with due consideration to the confounding effect of intramuscular fat required. The observations for CMT1A within the calf apply also to IBM, but at both thigh and calf level and with a greater magnitude of effect, as might be expected from knowledge of the disease distribution, pathogenesis and rate of progression (Table 4-8, Table 4-9, Table 4-10, Table 4-11). One specific example is that in calf muscles without significant fat infiltration, muscles which are STIR positive have T2 values of 41.8ms in controls, 44.5ms in CMT1A patients and 48.2ms in IBM patients. This greater degree of abnormal water distribution allows analysis of changes beyond those where there is not significant IFA. In Figure 4-8 we see the additional effect of STIR positivity on the T2 value beyond that due to the fat fraction measurement, with up to 25ms increase in T2 for STIR positive muscles versus STIR negative muscles in the same FF range. Whilst it is possible to calculate “T2 excess” through concurrent acquisition of T2 and FF data, the separation of fat and water effects on T2 and point of acquisition would be possible and is technically feasible, albeit with longer acquisition times (Janiczek et al., 2011).

#### **4.5.5 Challenges and study limitations**

There were a number of limitations in this study. We analysed data from only single slices of thigh and calf blocks, with only small ROI for T2 and MTR sequences. This may contribute variation if muscle pathology is anatomically non-homogenous. To ensure consistency in the volumes of tissue assessed longitudinally, slice positions were defined by measured distance from bony landmarks, a more reliable method than

surface-anatomy-based slice positioning (Fischmann et al., 2014), and follow up ROI drawn with direct reference to baseline ROI. T2 estimation by sampling two turbo-spin echo images is potentially less accurate than Carr-Purcell-Meiboom-Gill multiple-spin-echo methods. However, our approach provides time-efficient, wide-coverage quantification of T2 change using a method that can be implemented on standard MR systems without specialist modification.

#### **4.5.6 Future application**

Any proposed outcome measure must be quantitatively abnormal in patients compared with controls. For example mortality is not suitable as an outcome measure in diseases with normal life span such as CMT1A (Reilly et al., 2011) and IBM. The cross-sectional data have shown that quantifiable abnormalities of chronic pathology such as fat fraction and muscle atrophy would be applicable to studies requiring quantification of disease progression. In addition, measures of early abnormalities have been quantified with T2 and MTR differences in muscles without significant fat infiltration. These latter methods are sensitive to changes in water distribution which are reversible with effective therapy (Jurkat-Rott et al., 2009), so will be useful biomarkers in clinical trials enrolling patients with early or active disease.

Chronic IFA is common to a wide range of neuromuscular diseases. Although the precise molecular events responsible for IFA are not fully understood, one recent mechanistic study (Agle et al., 2013) suggests that a range of different primary genetic muscle diseases, and potentially denervation, stimulates myo-precursor stems cells to differentiate into adipose cells and fibroblasts. That IFA seems to be a common pathway in many genetic and acquired NMD contributes to the lack of diagnostic specificity of muscle MRI in NMD, but underlines its utility as an outcome measure that may be useful across NMD. As an example, we have shown here that the same types of changes can be quantified in these two very different diseases: an acquired late-onset progressive proximal and distal myopathy and an inherited childhood-onset slowly progressive distal predominant neuropathy.

Although differences between diseases result in different distributions and degrees of muscle abnormalities, their presence, direction and clinical correlations are consistent, making these outcome measures applicable across any neuromuscular disease with lower limb weakness. Through selection of MRI parameters targeted to disease and intervention, MRI outcome measures can be optimised to provide maximum responsiveness for a specific clinical trial. For example, the optimal MRI protocol for CMT1A would necessarily include assessment of fat infiltration of calf muscles with a sequence such as three-point Dixon. In IBM, to assess if an intervention reversed acute pathological processes, additional water-sensitive sequences such as T2 and

MTR quantification should be included. For any specific NMD, understanding of basic pathomechanisms, disease distribution and treatment mechanisms will enable trial-tailored selection of specific MRI outcome measures and appropriate anatomical imaging levels providing optimal responsiveness.

#### **4.6 Conclusions**

The second phase of this thesis study has examined cross-sectional MRI and clinical data and demonstrated excellent validity of quantitative MRI of the lower limb muscles in both CMT1A and IBM. T2 and MTR have additional sensitivity to disease processes prior to significant intramuscular fat accumulation. Beyond reliability and validity, which are found in other existing measures in neuromuscular diseases, outcome measure responsiveness is crucial to study power. This will be assessed using longitudinal data presented in chapter 5.

## 5 Longitudinal quantitative MRI in neuromuscular disease

### 5.1 *Introduction*

Whilst demonstration of reliability (chapter 3) and validity (chapter 4) is sufficient to determine that a health measurement is useful for descriptive or predictive purposes, knowledge of an outcome measure's responsiveness is needed before we can be confident about its application in a clinical trial (Kirshner and Guyatt, 1985). Responsiveness is often the domain where outcome measures in neuromuscular diseases are weakest (see 1.2.3) due to their slowly progressive nature. This is particularly so in CMT1A, where all outcome measures have shown low or moderate responsiveness (Pareyson et al., 2011a).

Responsiveness is simply defined as the ability of an instrument to detect when change has occurred (Roach, 2006). This can be divided into internal and external responsiveness. Internal responsiveness examines the change in the outcome measure over a specified time-frame, either by assessing the natural history of the disease as in this thesis, or assessing change in response to a known effective treatment (Roach, 2006). Comparing t-statistics for a known effective treatment is a preferred way to assess relative responsiveness of different measurements (Deyo et al., 1991; Roach, 2006), but this is not possible in diseases such as CMT1A and IBM where there are no known effective treatments. Alternative methods include looking at the "effect size", which is the mean change divided by the standard deviation of baseline scores (Deyo et al., 1991). Subsequently Guyatt (Guyatt et al., 1987) suggested the ratio of mean change to standard deviation of change (Equation 1-2, page 23), the responsiveness coefficient or standardised response mean. The standardised response mean is a key determinant of study power by Lehr's formula (Equation 1-3, page 23). The standardised response mean will be used as the main measure of responsiveness in this chapter.

These measures of responsiveness say nothing about whether the change is clinically meaningful: the "longitudinal validity" (Roach, 2006). One assessment of this is the external responsiveness, that is the degree to which changes in the measure correlate with changes in other measures of health status (Roach, 2006). It is worth noting that like validity and reliability, responsiveness is dependent on the specific group of patients being measured (Roach, 2006). It is not possible to assume that because an outcome measure is responsive in one disease, it will be equally responsive in another disease. Even within a single disease, responsiveness will vary for the same measurement depending on inclusion criteria. The SRM formula shows ways to

increase responsiveness through patient selection: by including those where the primary outcome measure is likely to change by the greatest magnitude and by the most uniform amount.

## **5.2 *Background literature***

There are significantly fewer longitudinal than cross-sectional quantitative MRI studies in neuromuscular diseases. As background, they will be briefly reviewed after reviewing all longitudinal qualitative studies in IBM and CMT.

### **5.2.1 Longitudinal qualitative MRI studies in CMT/IBM**

There are no longitudinal qualitative MRI natural history studies in IBM. In CMT1A, one recent study from Pelayo and colleagues examined T1w MRI, neurophysiology, muscle strength and other clinical measures in 14 patients with CMT1A over 2 years (Pelayo-Negro et al., 2014). Although many measurements showed cross-sectional correlation, none of the potential outcome measures including qualitative MRI showed significant change over two years.

### **5.2.2 Natural history quantitative MRI studies of muscle size**

Two natural history studies have looked at muscle size longitudinally, both in neuropathic conditions. Long-term natural history data are available in a small group of patients with diabetic neuropathy and controls. This showed a loss of muscle volume in ankle dorsiflexors and plantarflexors of 4.5% and 5.0% per year respectively in patients with diabetic neuropathy versus 1.7% and 1.8% annual loss in healthy controls (Andreassen et al., 2009). Standard deviations were not reported with small patient numbers but appeared to be approximately 1%, which would give an SRM of approximately 3 for the excess loss of muscle volume over one year in the patient group (highly responsive). This figure is inflated by the duration of follow-up assessment of 9-12 years after the original assessment, which would be impractical for a clinical trial, but there may be a good SRM over a shorter length of time. By contrast a study of patients with spinal muscular atrophy failed to show any change in thigh muscle volume over a six month interval (Sproule et al., 2011b). The study used a semi-automated segmentation method based on signal intensity of a T1w imaged and included 11 clinically heterogeneous subjects, which would have limited power to show significant change.

### **5.2.3 Interventional studies including muscle size measurements**

Perhaps surprisingly given the lack of natural history data there are a number of interventional studies using muscle size as an outcome measure in myopathic conditions. The earliest interventional study in a neuromuscular disease to include

quantitative MRI as an outcome measure was in fact performed in patients with IBM. In a small study of five patients with IBM undertaking a progressive resistance strength training programme for 12 weeks, Spector and colleagues report performing both lower and upper limb MRI at baseline and at study end, quantifying remaining muscle cross-sectional area based on signal intensities on a T1w sequence (Spector et al., 1997). The authors report that “significant changes in whole muscle cross-sectional area assessed by MRI were not seen in any of the muscles of patients as a result of training”, although no numerical values are given (neither cross-sectional nor longitudinal), nor details of which muscles were assessed, nor what change would be considered significant. Interestingly there were significant improvements in strength over the course of the study (Spector et al., 1997). If these two observations are accurate, it would imply an increase in the specific force of the muscle tissue (strength per cross-sectional area) indicating an improvement in muscle quality rather than quantity. However, the number of patients was very small, and with no control group the potential for significant systematic bias, such as with a learning effect, is high, so any conclusions from this study must be guarded.

Right thigh muscle volume measured on a proton density weighted MRI sequence was chosen as the primary outcome measure in a phase 2 clinical trial of bimagrumab in inclusion body myositis (Amato et al., 2014). Bimagrumab is an ActRII inhibitory antibody which blocks downstream signalling of TGF $\beta$ , a pathway which leads to muscle atrophy. A single dose was given to 11 patients whilst three patients received a placebo. Right thigh muscle volume increased by 6.5% compared with placebo ( $p=0.024$ ). Change in clinical outcome measures such as muscle strength and six minutes walking distance was more variable and generally not significant, as might be expected in a small short-duration trial. There was however a significant correlation between change in muscle volume and change in six minutes walking distance ( $r=0.61$ ,  $p=0.036$ ), demonstrating longitudinal validity. The standard deviation of the change in volume measurement was relatively high (approximately 4%), suggesting either poor reliability or significant variation in treatment response. Regardless, this study shows the potential power of volumetric assessment of muscle size if the intervention is expected to affect this parameter, though the functional benefit of simply increasing muscle size needs to be proven in a longer study.

Conversely a clinical trial of a myostatin inhibitor, also biologically predicted to increase muscle size, failed to show any significant effect on muscle volume using MRI in a mixed group of patients with muscular dystrophy. No significant change in lean body mass measured by DEXA was observed, suggesting the negative result may have been an ineffective drug rather than insensitive outcome measure. Good repeatability

of measurements was observed with standard deviation of change in volume measurements of approximately 1%.

MRI has also been used by an Italian group to investigate the effect of enzyme replacement therapy (ERT) in Pompe disease (Pichiecchio et al., 2009; Ravaglia et al., 2010), one of the few licensed drug therapies in an inherited neuromuscular disorder. They utilised segmentation of T1w lower limb MRI in 11 patients with Pompe disease receiving open label ERT for up to two years. They found a small increase in muscle volume of the less affected anterior compartment of the thigh (+7.5%) and an even larger increase in muscle strength (+45%), interpreted as an increase in both muscle size and quality (strength per unit volume) (Lynch et al., 1999), though some caution is needed in interpretation of the increased muscle strength due to the open-label uncontrolled design. The more affected posterior compartment of the thigh showed no increase in muscle volume. There was also overall increased intramuscular fat despite ERT. One explanation of these data is that there may be a “point of no return” in terms of intramuscular fat accumulation, where treatment is no longer effective, though this would need confirmation in larger blinded studies.

Considered together these trials highlight some of the advantages and disadvantages of primary use of muscle size parameters as outcome measures in neuromuscular diseases. The main advantage is that the images for analysis are relatively easy to obtain, as any sequence with adequate tissue contrast may be used. Most investigators have used T1w sequences to allow some estimation of the intramuscular fat accumulation using segmentation based on signal intensity, but this is a less precise means to quantify fat than a directly quantitative sequence such as three-point Dixon. The main disadvantages are the lack of natural history data to demonstrate that muscle atrophy is a primary marker of disease progression with the exception of diabetic neuropathy; the confounding effect of muscle atrophy associated with ageing; wide variation in muscle size between individuals; difficulties in precise measurement of muscle volume with relatively high standard deviations in longitudinal change observed; unclear relationship between muscle size and function; and an unclear role for drugs whose mechanism is not to directly promote muscle hypertrophy. Therefore, whilst muscle size may be helpful as a complementary measurement for some interventions in some diseases, it remains to be shown that it is universally applicable, indeed is likely not to be so for the reasons listed above.

#### **5.2.4 Natural history studies of quantitative tissue parameters**

The alternative to using muscle size outcome measures is utilising muscle tissue MRI parameters such as T2 relaxation time or fat fraction. An early study published in 1998 showed progression in T2 measurements from tibialis anterior muscle over a four

month interval in 11 patients with motor neuron disease (Bryan et al., 1998). Muscle size and T1 time did not change over this interval. Furthermore, T2 time correlated with motor amplitude on neurophysiology both cross-sectionally and longitudinally ( $r=-0.63$ ,  $p=0.037$ ). Unfortunately, the magnitude of change in T2 over 4 months was not reported, so responsiveness of this outcome measure cannot be calculated from the data provided.

Four more recent studies have measured quantitative muscle tissue MR parameters in muscular dystrophies: T2 in Duchenne (Willcocks et al., 2014), T2 and 2-point Dixon FF in oculopharyngeal muscular dystrophy (Fischmann et al., 2012), multi-spin echo FF in fascio-scapulo-humeral dystrophy (Janssen et al., 2014), and three-point Dixon FF in limb girdle muscular dystrophy type 2I including patients from our Centre (Willis et al., 2013). All four studies showed significant change in quantitative tissue parameters over the duration of the study.

In Duchenne, soleus whole tissue T2 was measured over 2 years in 15 ambulatory boys with Duchenne (Willcocks et al., 2014). This increased by  $5.5 \pm 6.6\%$  over 1 year (SRM 0.83) and  $9.5 \pm 7.4\%$  over 2 years (SRM 1.28). The absolute increase over 2 years was  $4.5 \pm 3.6\text{ms}$  (SRM 1.25), with significant increase in T2 also seen in peroneal muscles but not tibialis anterior. Change in soleus T2 was significantly correlated with change in three functional measurements undertaken ( $R=0.66-0.87$ ,  $p<0.01$ ), even though for only one of these (supine to stand time) was the change in the functional measure itself statistically significant. A limitation of the study was that only cross-sectional control data were collected. Fat-suppressed T2 values changed to a small degree over 2 years, suggesting most of the whole tissue T2 increase related to progression in intramuscular fat accumulation.

In OPMD, both T2 and 2-point Dixon FF were measured in all multiple thigh and calf muscles in 8 patients and 5 controls over a 13-month interval. Whilst functional measures showed no significant change over this interval, there was a 2.2ms increase in all muscle T2 and a 1.5% increase in fat fraction (both  $p<0.001$ ). Standard deviation was not reported directly, but estimated SRM from data depicted in figures was 1.1. FF and T2 were highly correlated. These data demonstrate that MRI can detect subclinical disease progression over a short time period with high responsiveness.

In FSHD, 41 patients were imaged with 11 of these reimaged at a 4 month interval (Janssen et al., 2014). A single thigh was imaged using both qualitative techniques and a fat fraction estimation based on multi-spine echo sequence (in muscles without oedema). The overall increase in fat fraction was  $1.8 \pm 4\%$  (SRM =0.45), although this was greater when only muscles with intermediate baseline fat-fraction between 25 and 75% were considered ( $6 \pm 5\%$ ). The inclusion of multiple muscles from individual

patients within the statistical calculation, the different technique needed for muscles with oedema and short follow up were limitations of this study.

Finally, in LGMD2I, 32 patients across 4 international sites were studied using three-point Dixon fat fraction quantification to quantify thigh and calf muscle fat fraction of a single lower limb over a 12 month interval. Significant increases in fat fractions were seen in 9/14 muscles examined, but no significant change in functional measures. Increases ranged from 2% in biceps femoris to 0% in tibialis anterior. This study again demonstrated superior responsiveness of MRI over functional outcome measures.

### **5.2.5 Interventional studies including quantitative muscle tissue parameters**

There are three intervention studies with quantitative muscle tissue MR parameter measures included as outcome measures: two with steroids in Duchenne (Arpan et al., 2014; Kim et al., 2010b) and one with exercise in juvenile dermatomyositis (Maillard et al., 2005). The study in juvenile dermatomyositis utilised T2 quantification as a safety parameter with measurements before, and 30 and 60 minutes after exercise in a physiotherapy-led exercise programme (Maillard et al., 2005). No significant change was seen in MRI, muscle strength or blood parameters in 20 children in juvenile dermatomyositis or 20 controls, suggesting the exercise programme did not induce muscle oedema.

Corticosteroid treatment is recognised as an effective treatment in Duchenne muscular dystrophy. Benefits include prolonged ambulation, improved muscle strength, improved pulmonary function, delayed cardiomyopathy onset and reduced need for scoliosis surgery (Gloss et al., 2016). It is therefore a logical intervention to assess using MRI, although its proven efficacy means a randomized placebo controlled trial design would be unethical. The first study performed T2 mapping of right gluteus maximus muscle in 11 boys with DMD before and after commencing corticosteroid treatment with a mean duration of 14.6 months (Kim et al., 2010b). Qualitative T1w and water sensitive sequences were also performed. The changes were variable: 2 boys had increases in fat infiltration on conventional imaging associated with an increase in T2. Of the 9 who showed no change on qualitative imaging, three showed an increase in T2, two showed no change in T2 and four showed a reduction in T2 within gluteus maximus muscle. Since T2 time is affected by both changes in both water and fat distribution within muscle, the changes may relate to alterations in either, and the study emphasises the need for multi-parameter quantitative MRI assessment to allow disambiguation of effects. The study population was quite heterogeneous with an age

range of 5-14 years, so without a control group it is difficult to separate the effect of the corticosteroid treatment from natural disease processes.

The later study overcame these limitations by including both fat fraction and T2 time, assessing a homogenous patient group (aged 5-6.9 years), and longitudinal assessment of patients who did and did not start steroid therapy, although treatment assignment was not randomized (Arpan et al., 2014). Cross-sectional analysis included 15 boys on corticosteroids and 15 corticosteroid naïve patients and showed significantly increased T2 and fat-fraction in the latter group. Six boys who remained off corticosteroids had a significant increase in fat fraction over 12 months (+7% in vastus lateralis and +3% in soleus), which was significantly greater than 9 boys who remained on corticosteroids (+1% in vastus lateralis and +0% in soleus). Finally, previously corticosteroid naïve boys who started treatment had a significant decrease in muscle T2 at both 3 and 6 months without a change in measured fat fraction. Although observational rather than randomised, these data demonstrate the potential of MRI to quantify a reduction in muscle oedema over a short time interval and a reduction in progressive intramuscular fat accumulation over a longer time interval in response to an intervention.

### **5.2.6 Summary**

Over recent years there is a growing body of longitudinal data on quantitative MRI in neuromuscular diseases. Overall quantitative muscle tissue parameters appear more responsive than size-based measurements, particularly in primary muscle disorders, though the inclusion of size-based measurements is attractive, particular if the intervention is expected to have a biological effect on muscle size. The importance of multi-parameter assessment in interventional studies is emphasised to allow separation of the effects of treatment on muscle water distribution in the short term and intramuscular fat accumulation in the longer term. Studies consistently show MRI outcome measures to be more responsive than clinical outcome measures, though some studies are able to show longitudinal correlation between outcome measures. The overall scope of the studies in terms of range of diseases investigated and patient numbers is relatively small.

### **5.3 *Longitudinal study aims***

The primary aim of the longitudinal study phase is to assess internal responsiveness by calculating standardised response means for quantitative MRI parameters and comparing these with those of clinical and myometric measures. External responsiveness will be assessed by correlating significant changes in MRI and clinical or myometric parameters. Finally, any significant predictors of change will be determined as a means of optimising responsiveness in future trials.

## **5.4 Results**

### **5.4.1 Study subjects**

Repeat assessments were performed after a mean interval of 12.4 months (s.d. 1.0 months) in 62/69 (90%) subjects (CMT1A: 18/20, 1 was withdrawn due to ankle surgery, 1 withdrew from the study; IBM: 18/20, 2 withdrew as no longer able to travel; controls: 26/29, 3 had attended with patients who withdrew). Figure 2-2 (page 41) shows a flow chart of participants through the study.

### **5.4.2 Overall longitudinal data**

Longitudinal change of clinical, myometric and qualitative and quantitative MRI outcome measures are summarised in Table 5-1 below. Change in patients was considered both whether it was significantly different from 0 (paired t-test, p1 in table) and whether it was significantly different to that seen in matched controls (student two-tailed t-test, p2 in table). In CMT1A patients, only change in whole ROI calf FF was significant in both tests, whilst in IBM change was significant in many quantitative parameters.

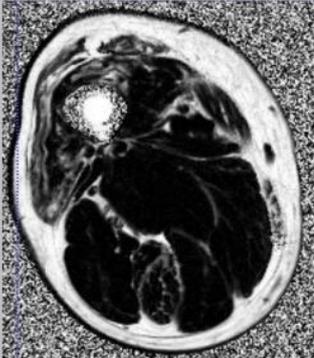
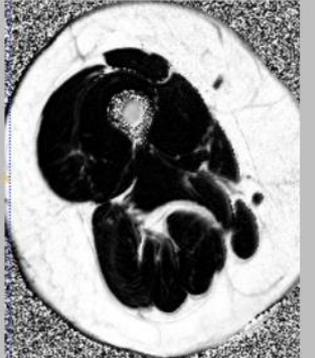
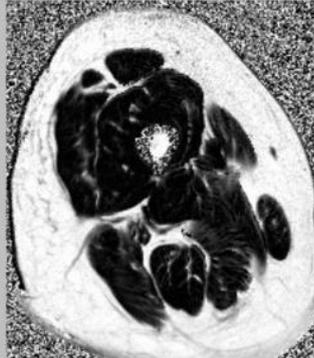
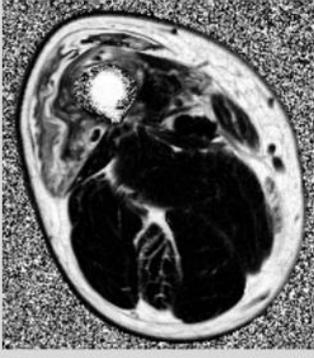
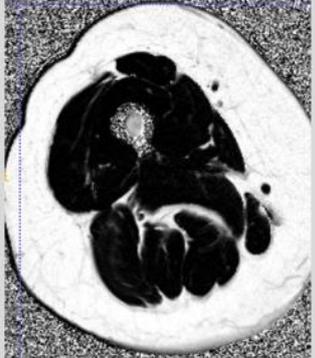
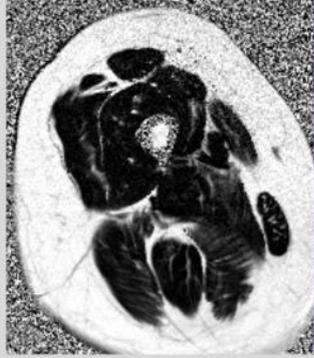
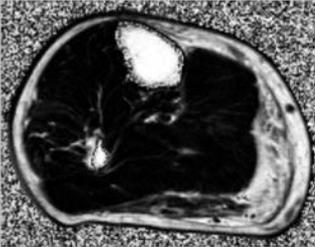
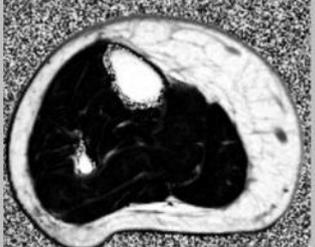
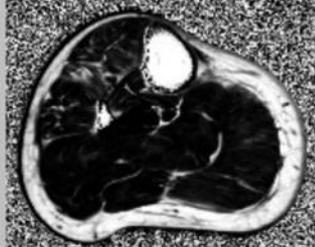
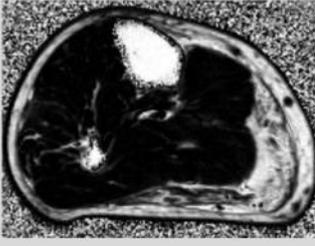
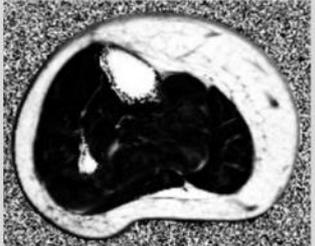
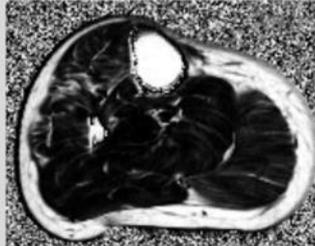
Table 5-1: Overall longitudinal data

Measure	CMT	Control (CMT)	p1	p2	IBM	Control (IBM)	p1	p2
<b>Thigh MRI</b>								
FF (%)	0.4 ± 1.0	0.0 ± 0.3	0.15	0.12	3.3 ± 4.0	-0.1 ± 0.4	0.005	<0.001
T1 (ms)	23 ± 39	14 ± 42	0.03	0.52	-44 ± 44	19 ± 44	0.002	<0.001
T2 (ms)	1.3 ± 1.5	0.6 ± 1.6	0.003	0.21	2.6 ± 4.2	0.5 ± 1.6	0.03	0.07
MTR (p.u.)	0.0 ± 0.7	-0.1 ± 0.3	0.96	0.55	-0.9 ± 1.6	0.0 ± 0.5	0.06	0.06
FF whole (%)	0.2 ± 0.8	0.2 ± 0.8	0.38	0.97	3.3 ± 3.2	0.2 ± 0.8	<0.001	<0.001
CSA (% change)	-0.6 ± 5.3	-0.7 ± 7.2	0.62	0.96	-2.7 ± 7.9	0.2 ± 5.7	0.08	0.23
Mercuri grade	-0.02 ± 0.21	0.01 ± 0.21	0.72	0.72	0.00 ± 0.30	-0.01 ± 0.26	0.97	0.91
STIR	0.09 ± 0.23	0.05 ± 0.07	0.45	0.45	0.25 ± 0.31	0.05 ± 0.05	0.005	0.01
<b>Thigh Myometry</b>								
Knee ext. (Nm)	1.0 ± 8.1	-5.2 ± 10.2	0.57	0.06	-6.0 ± 5.2	-4.2 ± 11.4	<0.001	0.55
Knee flex. (Nm)	2.5 ± 6.8	1.7 ± 7.6	0.15	0.75	-1.7 ± 4.3	2.9 ± 7.1	0.08	0.02
<b>Calf MRI</b>								
FF (%)	1.1 ± 2.4	0.0 ± 0.4	0.07	0.07	2.6 ± 2.7	0.0 ± 0.4	0.004	<0.001
T1 (ms)	22 ± 44	37 ± 43	0.14	0.40	1 ± 55	23 ± 44	0.94	0.27
T2 (ms)	1.4 ± 2.6	0.3 ± 1.1	0.05	0.13	4.5 ± 3.7	0.0 ± 1.5	<0.001	<0.001
MTR (p.u.)	-0.2 ± 0.6	0.0 ± 0.7	0.30	0.28	-0.7 ± 0.7	0.2 ± 0.8	0.004	0.003
FF whole (%)	1.2 ± 1.5	0.2 ± 0.4	0.002	0.008	2.6 ± 2.4	0.1 ± 0.4	0.002	<0.001
CSA (% change)	0.6 ± 6.6	0.0 ± 4.4	0.67	0.74	-2.5 ± 3.9	0.1 ± 5.0	0.01	0.11
Mercuri grade	0.01 ± 0.44	0.06 ± 0.12	0.92	0.68	0.18 ± 0.29	0.07 ± 0.09	0.03	0.17
STIR	0.14 ± 0.28	0.09 ± 0.20	0.05	0.53	0.27 ± 0.32	0.05 ± 0.18	0.003	0.02
<b>Calf Myometry</b>								
Ankle PF (Nm)	3.8 ± 7.6	2.1 ± 10.8	0.05	0.59	-0.7 ± 3.9	5.6 ± 14.6	0.42	0.09
Ankle DF (Nm)	2.1 ± 4.0	0.5 ± 3.8	0.06	0.24	-0.4 ± 4.2	2 ± 5.3	0.65	0.14
Ankle Inversion	1.8 ± 5.4	0.4 ± 5.0	0.18	0.45	-1.5 ± 2.5	0.3 ± 4.5	0.02	0.14
Ankle Eversion	1.4 ± 2.1	0.9 ± 2.3	0.02	0.53	-0.1 ± 1.7	0.5 ± 2.4	0.77	0.40
<b>Overall Clinical measures</b>								
Lower Limb MRC	-0.4 ± 3.8	NA	0.65	NA	-3.4 ± 5.6	NA	0.02	NA
SF-36 (%)	-2.5 ± 15.2	NA	0.51	NA	-1.1 ± 7.0	NA	0.55	NA
SF-36 PF (%)	-0.9 ± 12.3	NA	0.77	NA	-4.1 ± 18.5	NA	0.39	NA
CMTES (points)	-0.3 ± 1.3	NA	0.37	NA	NA	NA	NA	NA
IBM FRS (points)	NA	NA	NA	NA	-2.8 ± 2.9	NA	<0.001	NA

p1: p-value of paired t-test in patient group; p2: p-value of two-tailed t-test patient vs matched control; CMT: Charcot-Marie-Tooth disease; IBM: inclusion body myositis; FF: fat fraction; MTR: magnetisation transfer ratio; MRC: Medical Research Council bedside strength assessment; SF-36: Short Form Health Survey; PF: physical functioning domain; CMTES: CMT Examination Score; IBM-FRS: IBM Functional Rating Score; NA: Not applicable. Highlighted measures are those where follow-up value is significantly different from baseline AND where applicable change seen is significantly different from change in controls (both paired t-test and two-tailed t-test are significant).

### 5.4.3 Longitudinal assessment of fat atrophy

Figure 5-1: Longitudinal fat fraction images

Thigh Dixon	IBM	Volunteer	CMT
Baseline	 18.8% Fat	 6.0% Fat	 13.7% Fat
One year	 25.0% Fat	 7.0% Fat	 14.9% Fat
Calf Dixon	IBM	Volunteer	CMT
Baseline	 19.9% Fat	 4.7% Fat	 9.6% Fat
One year	 20.5% Fat	 3.7% Fat	 12.5% Fat

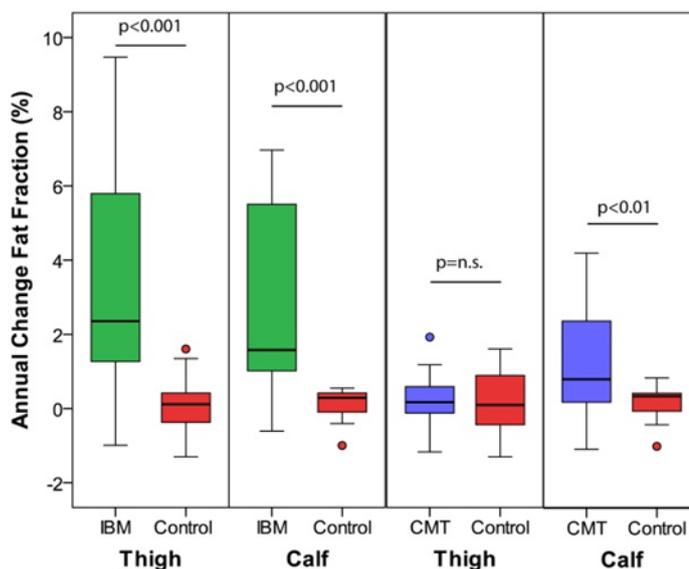
Sample longitudinal fat fraction images for a subject in each group. Mean whole muscle FF for each slice is given below each image. Greatest change in seen at thigh level in IBM and calf level in CMT1A.

### 5.4.3.1 Qualitative imaging

Changes in T1w and STIR qualitative imaging scores were not significant in CMT1A patients either with paired t-test or compared with matched controls, who as expected also showed no significant change. In T1w analysis of IBM patients there was a small increase in mean Mercuri grade at calf level over 12 months (+0.18,  $p=0.03$ ) but this was not significant when compared with a small (non-significant) increase seen in controls (student t-test  $p=0.17$ ). By contrast a significant increase in mean STIR grading was noted for IBM patients at both thigh and calf level, suggesting an increase in muscles with abnormal muscle water distribution over one year.

### 5.4.3.2 Quantitative imaging

Figure 5-2: Box and whisker plot of longitudinal change in fat fraction



Group comparison against matched controls shows that significant increases in overall mean fat fraction are seen in IBM patients at thigh and calf level ( $p < 0.001$ ) and at calf level in CMT1A patients ( $p < 0.01$ ). Bars indicate median, 25<sup>th</sup>, 50<sup>th</sup> and 75<sup>th</sup> centiles, lines: range, °: minor outlier

The overall change in FF over 12 months is shown in Figure 5-2. Significant change in FF is seen at both thigh and calf level in IBM patients, and at calf level in CMT1A patients. As all quantitative parameters in this study exhibit strong interdependencies (see Table 3-14 and Table 4-7) due to their common sensitivity to changes in intramuscular fat accumulation, similar magnitude of change might be expected across these measurements with increases in FF and T2, and reductions in T1 and MTR in circumstances of progression.

#### 5.4.3.2.1 Controls

With regard to the controls, there was no change in FF, T2 or MTR over the one year follow-up, as would be expected from the small magnitude of age dependency seen in

the cross-sectional volunteer data (see Table 3-12). There was however a small but significant increase in measured T1 time at both thigh and calf level, suggesting a systematic bias in either acquisition or analysis between the two time points. Unlike with T2 (see 2.4.4.3.1), this difference did not clearly relate to the scanner software upgrade.

#### **5.4.3.2.2 CMT1A**

In the CMT1A group, at thigh level there were no changes in quantitative parameters significant in both paired-t-test and compared with controls by student t-test. Of note, there was a significant increase in T1 time compared with baseline ( $+23\pm 39\text{ms}$ ,  $p=0.03$ ), which could indicate a reduction in IFA, but with a near identical change seen in the healthy controls, systematic bias is the more likely explanation. The overall change in FF at thigh level was not significant, though baseline pathology was rare here as would be expected from the distal predominant phenotype. In the three CMT1A patients with thigh fat fraction over 5% at baseline, there was an average 0.93% increase in thigh fat fraction over 12 months, although the number here is clearly too small to allow any test of statistical significance. There was no significant change in thigh muscle cross-sectional area.

By contrast at calf level there was significant increase in whole ROI FF in CMT1A patients:  $1.2 \pm 1.5\%$  ( $p=0.002$  paired t-test,  $p=0.008$  vs controls). There was a similar magnitude of change ( $+1.1\%$ ) in small ROI FF, but greater variability (s.d. 2.4%), presumably due to lesser reliability, meant this change failed to reach statistical significance ( $p=0.07$ ). Similarly mean calf T2 was greater and mean calf MTR was smaller at follow-up than baseline, as would be expected from their common sensitivity to intramuscular fat accumulation, but again the change was more variable so did not approach statistical significance. Calf muscle CSA was not significantly changed.

#### **5.4.3.2.3 IBM**

In IBM patients at 12-month follow up, all-muscle thigh- and calf-level FF for both small and whole muscle ROIs, and calf muscle T2 had increased significantly from baseline, whereas calf muscle MTR significantly decreased (Table 5-1). Thigh level increases in T2 ( $p=0.03$ ,  $p=0.07$ ) and reductions in MTR ( $p=0.06$ ,  $p=0.06$ ) did not reach statistical significance for both paired-t-test and two-tailed t-test versus controls. Consistent with observation in controls and CMT1A patients, T1 measurements were most variable, with a significant decrease at thigh level, but no change at calf level. These data reflect a significant progression in IFA at both thigh and calf level over 12 months, with three-point Dixon FF the most reliable and sensitive means of quantifying this. There was also a progression of muscle atrophy at thigh level with all-muscle CSA decreasing versus baseline ( $p<0.05$ ) at calf but not at thigh level ( $p=0.08$ ).

The quadriceps and hamstrings reduced in CSA by a similar percentage (quadriceps:  $-2.9 \pm 7.6\%$ ; hamstrings  $-3.3 \pm 11.3\%$ ), and increased in FF by a similar amount (quadriceps:  $4.2 \pm 5.7\%$ ; hamstrings:  $2.7 \pm 2.0\%$ ;  $p=0.33$ ). Individual patterns of progression varied with FF changes in quadriceps greater in early disease, and changes in hamstrings CSA greater in late disease, consistent with the clinically recognised progression of muscle involvement.

#### **5.4.4 Longitudinal clinical data**

##### **5.4.4.1 CMT1A**

In the CMT1A group no clinical measure showed significant change over 12 months. Myometry measured ankle dorsiflexion and eversion strength showed an apparent increase at follow-up, though this was also seen in the CMT matched volunteers so may represent a learning effect.

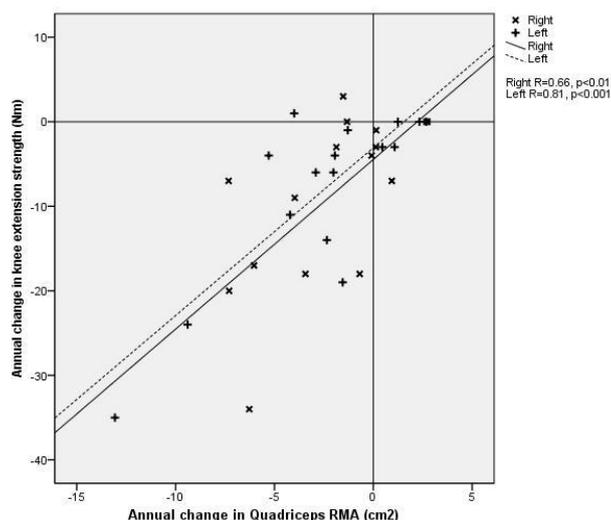
##### **5.4.4.2 IBM**

Of the clinical parameters (Table 5-1), significant change was seen in the IBM patient group in MRC scores and IBM-FRS. IBM patients showed a significant reduction in overall knee extension ( $p<0.001$ ) compared with baseline and a trend to reduction in knee flexion strength ( $p=0.08$ ). However, the reduction in knee extension strength was not significantly different to age matched controls ( $p=0.55$ ), whilst the reduction in knee flexion strength was ( $p=0.02$ ). In all groups, the standard deviations of change in muscle strength measurements were relatively high, suggesting overall poor reliability of these measurements.

#### **5.4.5 Longitudinal MRI - clinical correlation**

Correlations between change in clinical parameters and change in MRI parameters were not significant in the CMT group. However, in IBM the change in myometric strength of knee extension correlated with change in several MRI parameters including quadriceps remaining muscle area (right  $R=0.66$ ,  $p=0.005$ , left  $R=0.81$ ,  $p=0.0001$ , Figure 5-3).

Figure 5-3: Correlation of change in strength and remaining muscle area on MRI



Change in the MRI metric of quadriceps remaining muscle area (RMA) correlated with change in quadriceps strength over 12 months in IBM patients on both left and right thighs

#### 5.4.6 Predictors of change

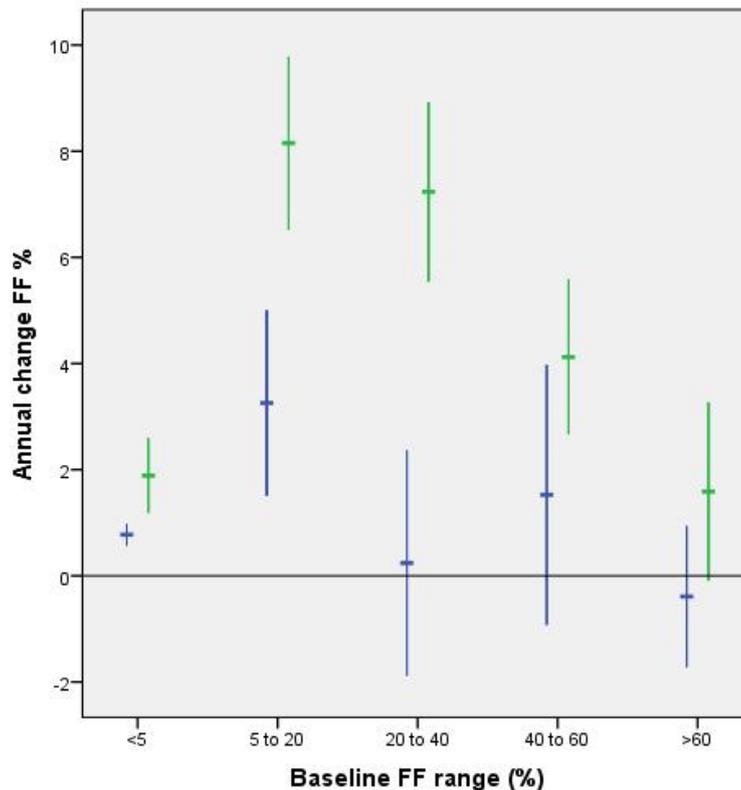
Annual change in fat fraction was found to relate to both baseline fat fraction and presence of STIR hyperintensity (Table 5-2). Greatest change was seen when fat infiltration was moderate (20-60%) and STIR hyperintensity was marked. Note that in the control group a small reduction in FF in those with baseline fat fraction greater than 5% is likely due to a regression to the mean effect. Data just for IBM thigh muscle are shown in Figure 5-4. Across the range of baseline fat-fractions, progression was greater in muscles with STIR hyperintensity. This difference was statistically significant for the <5%, 5-20% and 20-40% bands of baseline fat fraction. The significant difference in muscles with “normal” (less than 5%) fat fraction at baseline suggests that STIR hyperintensity precedes fat accumulation in early disease.

Table 5-2: Annual change in fat fraction within muscles varies with baseline characteristics

Sequence	Category	CMT	IBM	Controls
STIR hyperintensity	Absent	0.5 ± 0.1	1.1 ± 0.3	0.0 ± 0.0
	Mild	1.8 ± 0.7*	3.9 ± 0.6*	0.0 ± 0.1
	Marked	2.4 ± 1.4*	5.9 ± 1.0*	0.4 ± 0.9
Fat fraction	0 to 5	0.3 ± 0.1	1.0 ± 0.2	0.1 ± 0.0
	5 to 20	0.3 ± 0.5	6.2 ± 1.0*	-1.0 ± 0.5*
	20 to 40	4.6 ± 2.2*	6.9 ± 1.3*	ND
	40 to 60	7.3 ± 2.2*	3.3 ± 1.0	ND
	>60	1.6 ± 1.1	1.0 ± 0.9	ND

Data are mean ± SEM, thigh and calf muscles combined, \* statistically significant p<0.05 ANOVA/Tukey; CMT: Charcot-Marie-Tooth disease; IBM: inclusion body myositis; STIR: short-tau-inversion-recovery; ND: no data points.

Figure 5-4: Longitudinal change in thigh muscles in IBM patients grouped by baseline FF and STIR



Annual change in fat fraction within thigh muscles in IBM patients varies depending on baseline fat fraction range and baseline STIR appearance, with greatest rate of change seen where baseline STIR hyperintensity is present and fat baseline fraction is moderate (5-40%). Green: STIR hyperintensity present; blue: STIR hyperintensity absent. Values are mean  $\pm$  1 standard error.

#### 5.4.6.1 Baseline T2 as predictor of change

As is noted above, the predominant effect of fat infiltration on T2 and MTR measurements means absolute values of T2 and MTR are not reliable markers of changes in muscle water distribution. However due to the concurrent measurement of fat fraction within the same volume of tissue, it was possible to model the relationship between T2 and FF to calculate a “T2 excess” as a surrogate measure of muscle water T2. This is defined as the difference in measured T2 and predicted T2 based on FF measurement. Fit lines were drawn on the basis that changes in muscle water distribution tend to increase T2 rather than decrease T2, which was evident on a scatter plot of the T2/FF relationship (Figure 5-5). Baseline T2 excess values were then grouped as less than 5ms, between 5 and 15ms and >15ms, as being roughly equivalent to the three qualitative STIR gradings, and subsequent change in FF calculated (Table 5-3). At both thigh and calf level FF change was highest in muscle with greatest T2 excess.

Figure 5-5: Correlation between T2 and FF at baseline to estimate "baseline T2 excess"

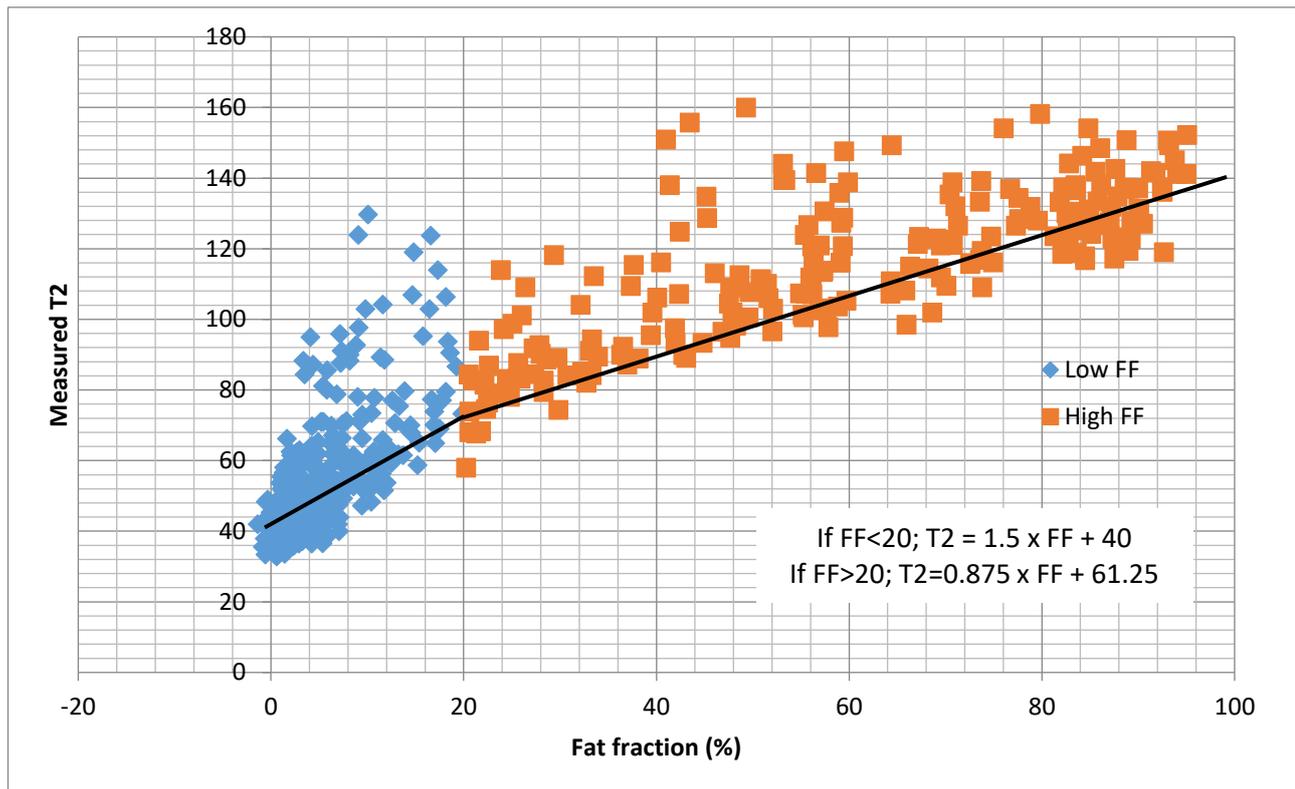


Table 5-3: Change in FF grouped by "T2-excess" at baseline

Baseline T2 excess (ms)	<5	5 to 15	>15
FF change Thigh (%)	2.33	3.72	4.51
FF change Calf (%)	2.29	3.18	4.10

#### 5.4.7 Standardised Response Mean

In the IBM group, significant 12-month change was observed in clinical, myometric and MRI measures. The highest SRMs, greater than 1, were for whole muscle FF at thigh and calf level, and T2 at calf level, whilst no clinical parameters had SRM greater than 1 (Table 5-4). In the CMT group, significant 12-month change occurred only in whole muscle FF at calf level, with an SRM of 0.83 (Table 5-5).

Table 5-4: Standardised Response Means in IBM

Measure	IBM Mean±s.d. (95% CI)	Control (IBM) Mean±s.d. (95% CI)	p vs baseline	p vs control	SRM
<b>Overall Clinical measures</b>					
<b>Lower Limb MRC</b>	-3.4 ± 5.6 (0.6;6.2)	NA	0.02	NA	0.63
<b>SF-36 (%)</b>	-1.1 ± 7.0 (-2.7;4.8)	NA	0.55	NA	0.16
<b>SF-36 PF (%)</b>	-4.1 ± 18.5 (-13.9;5.8)	NA	0.39	NA	0.22
<b>IBM FRS (points)</b>	-2.8 ± 2.9(1.3;4.2)	NA	<0.001	NA	0.97
<b>Thigh Myometry</b>					
<b>Knee ext. (Nm)</b>	-6.0 ± 5.2 (-8.4;-3.6)	-4.2 ± 11.4 (-9.5;1.1)	<0.001	0.55	-1.15
<b>Knee flex. (Nm)</b>	-1.7 ± 4.3 (-3.7;0.2)	2.9 ± 7.1 (-0.4;6.2)	0.08	0.02	-0.40
<b>Calf Myometry</b>					
<b>Ankle PF (Nm)</b>	-0.7 ± 3.9 (-2.5;1.1)	5.6 ± 14.6 (-1.2;12.4)	0.42	0.09	-0.19
<b>Ankle DF (Nm)</b>	-0.4 ± 4.2 (-2.4;1.5)	2.0 ± 5.3 (-0.4;4.4)	0.65	0.14	-0.10
<b>Thigh MRI</b>					
<b>FF (%)</b>	3.3 ± 4.0 (1.4;5.2)	-0.1 ± 0.4 (-0.3;0.1)	0.005	<0.001	0.83
<b>T2 (ms)</b>	2.6 ± 4.2 (0.5;4.7)	0.5 ± 1.6 (-0.3;1.3)	0.03	0.07	0.62
<b>MTR (p.u.)</b>	-0.9 ± 1.6 (-1.7;0.0)	0.0 ± 0.5 (-0.2;0.2)	0.06	0.06	-0.54
<b>FF whole (%)</b>	3.3 ± 3.2 (1.8;4.9)	0.2 ± 0.8 (-0.2;0.6)	<0.001	<0.001	1.06
<b>CSA (% change)</b>	-2.7 ± 7.9 (-6.5;1.1)	0.2 ± 5.7 (-2.5;2.9)	0.08	0.23	-0.34
<b>Calf MRI</b>					
<b>FF (%)</b>	2.6 ± 2.7 (1.1;4.1)	0.0 ± 0.4 (-0.2;0.2)	0.004	<0.001	0.97
<b>T2 (ms)</b>	4.5 ± 3.7 (2.6;6.4)	0.0 ± 1.5 (-0.7;0.7)	<0.001	<0.001	1.21
<b>MTR (p.u.)</b>	-0.7 ± 0.7 (-1.1;-0.3)	0.2 ± 0.8 (-0.2;0.6)	0.004	0.003	-0.99
<b>FF whole (%)</b>	2.6 ± 2.4 (1.3;4.0)	0.1 ± 0.4 (-0.1;0.3)	0.002	<0.001	1.07
<b>CSA (% change)</b>	-2.5 ± 3.9 (-4.4;-0.6)	0.1 ± 5.0 (-2.3;2.5)	0.01	0.11	-0.63

*p vs baseline: p-value of paired t-test in patient group; p vs control: p-value of two-tailed t-test patient vs matched control; CI: confidence interval for mean change; IBM: inclusion body myositis; FF: fat fraction; MTR: magnetisation transfer ratio; MRC: Medical Research Council bedside strength assessment; SF-36: Short Form Health Survey; PF: physical functioning domain; IBM-FRS: IBM Functional Rating Score; NA: Not applicable. Highlighted measures are those where follow-up value is significantly different from baseline AND where applicable change seen is significantly different from change in controls (both paired t-test and two-tailed t-test are significant).*

Table 5-5: Standardised Response Means in CMT1A

Measure	CMT Mean±s.d. (95% CI)	Control (CMT) Mean±s.d. (95% CI)	p vs baseline	p vs control	SRM
<b>Overall Clinical measures</b>					
<b>Lower Limb MRC</b>	-0.4 ± 3.8 (-1.5;2.3)	NA	0.65	NA	-0.11
<b>SF-36 (%)</b>	-2.5 ± 15.2 (-5.3;10.3)	NA	0.51	NA	-0.16
<b>SF-36 PF (%)</b>	-0.9 ± 12.3 (-5.4;7.2)	NA	0.77	NA	-0.08
<b>CMTES (points)</b>	-0.3 ± 1.3 (-0.9;0.4)	NA	0.37	NA	-0.23
<b>Thigh Myometry</b>					
<b>Knee ext. (Nm)</b>	1.0 ± 8.1 (-2.8;4.8)	-5.2 ± 10.2 (-9.9;-0.4)	0.57	0.06	0.12
<b>Knee flex. (Nm)</b>	2.5 ± 6.8 (-0.7;5.8)	1.7 ± 7.6 (-1.7;5.2)	0.15	0.75	0.37
<b>Calf Myometry</b>					
<b>Ankle PF (Nm)</b>	3.8 ± 7.6 (0.0;7.4)	2.1 ± 10.8 (-2.9;7.1)	0.06	0.59	0.51
<b>Ankle DF (Nm)</b>	2.1 ± 4.0 (0.2;4.0)	0.5 ± 3.8 (-1.3;2.2)	0.05	0.24	0.51
<b>Thigh MRI</b>					
<b>FF (%)</b>	0.4 ± 1.0 (-0.1;0.9)	0.0 ± 0.3 (-0.1;0.1)	0.15	0.12	0.36
<b>T2 (ms)</b>	1.3 ± 1.5 (0.6;2.1)	0.6 ± 1.6 (-0.2;1.4)	0.003	0.21	0.86
<b>MTR (p.u.)</b>	0.0 ± 0.7 (-0.3;0.3)	-0.1 ± 0.3 (-0.3;0.1)	0.96	0.55	-0.01
<b>FF whole (%)</b>	0.2 ± 0.8 (-0.2;0.6)	0.2 ± 0.8 (-0.2;0.6)	0.38	0.97	0.22
<b>CSA (% change)</b>	-0.6 ± 5.3 (-3.1;1.9)	-0.7 ± 7.2 (-4.0;2.6)	0.62	0.96	-0.12
<b>Calf MRI</b>					
<b>FF (%)</b>	1.1 ± 2.4 (0.2;2.2)	0.0 ± 0.4 (-0.2;0.2)	0.07	0.07	0.46
<b>T2 (ms)</b>	1.4 ± 2.6 (0.0;2.6)	0.3 ± 1.1 (-0.2;0.9)	0.05	0.13	0.54
<b>MTR (p.u.)</b>	-0.2 ± 0.6 (-0.5;0.1)	0.0 ± 0.7 (-0.3;0.4)	0.30	0.28	-0.34
<b>FF whole (%)</b>	1.2 ± 1.5 (0.5;1.9)	0.2 ± 0.4 (0.0;0.4)	0.002	0.008	0.83
<b>CSA (% change)</b>	0.6 ± 6.6 (-2.5;3.8)	0.0 ± 4.4 (-2.0;2.0)	0.67	0.74	0.10

*p vs baseline: p-value of paired t-test in patient group; p vs control: p-value of two-tailed t-test patient vs matched control; CI: confidence interval for mean change; CMT: Charcot-Marie-Tooth disease; FF: fat fraction; MTR: magnetisation transfer ratio; MRC: Medical Research Council bedside strength assessment; SF-36: Short Form Health Survey; PF: physical functioning domain; CMTES: CMT Examination Score; NA: Not applicable. Highlighted measures are those where follow-up value is significantly different from baseline AND where applicable change seen is significantly different from change in controls (both paired t-test and two-tailed t-test are significant)*

## **5.5 Discussion**

Significant changes in many MRI measurements over 12 months were seen in the IBM patients, with correlation between change in strength and change in MRI measured remaining muscle area providing evidence for longitudinal validity. In CMT1A, only whole ROI fat fraction at calf level showed significant change over 12 months with large responsiveness (SRM 0.83). Fat fraction progression was greatest in muscles with moderate baseline fat fraction and/or markers of abnormal water distribution, providing insights into disease pathogenesis and means to increase responsiveness.

### **5.5.1 Outcome measure responsiveness**

Outcome measure sensitivity to change can be expressed by the standardised response mean (SRM), calculated as the ratio of mean change to standard deviation of change (Liang et al., 1990), categorised by magnitude according to Cohen's rule of thumb: < 0.2 minimal responsiveness; 0.2-0.5 small responsiveness; 0.5-0.8 moderate responsiveness; >0.8 large responsiveness. Responsiveness is a key determinant of study power, with an inverse square relation to sample size by Lehr's formula (Lehr, 1992), such that doubling outcome measure responsiveness results in a four-fold decrease in sample size required for equivalent study power.

#### **5.5.1.1 Responsiveness of outcome measures in CMT1A**

The development of responsive outcome measures has proven especially difficult to date in NMD such as CMT1A. The CMTNS was selected as the most appropriate primary outcome measure for CMT trials at the 136th ENMC workshops and was used in the ascorbic acid trials in adults with CMT1A (Lewis et al., 2013; Micallef et al., 2009; Pareyson et al., 2011a). The calculated SRM of the CMTNS and other outcome measures in the largest of these (Pareyson et al., 2011a), together with five year natural history data (Verhamme et al., 2009), are shown in table 4. On average, a 0.3points per year CMTNS increase is observed, resulting in minimal responsiveness over two years and small responsiveness over five years. Neurophysiology either showed no significant change from baseline (Pareyson et al., 2011a) or change no greater than seen in controls (Verhamme et al., 2009), so appears unsuitable as an outcome measure. Ankle dorsiflexion showed moderate responsiveness over two years using a custom-built frame but no responsiveness over five years with hand-held myometry. Timed functional tests and pinch or grip myometry showed small responsiveness over both two and five years. The responsiveness of myometry, MRC strength assessment, CMTES and SF36-PF in this current study are consistent with those data (Table 5-6). In marked contrast, in this thesis, calf-level MRI-quantified FF

showed large responsiveness (SRM=0.83) with a highly significant ( $p=0.002$ ) increase in FF at 12 months.

This has very important implications for future trial design. For a hypothetical CMT1A treatment trial powered to detect a 50% reduction in disease progression over one year with 80% power at  $p<0.05$  significance, the number of patients needed in active and placebo arms (Lehr, 1992) would be approximately 93 using calf-muscle MRI-determined FF as the primary outcome measure, as opposed to 7700 patients for equivalent statistical power using CMTNS to quantify outcome. MRI-quantified calf muscle FF is therefore the most responsive outcome measure proposed to date in CMT1A.

Table 5-6: Comparison of outcome measure responsiveness with prior CMT1A studies

Measure	Baseline	Change	p	SRM
<i>This study, 1 year</i>				
<b>Mean calf fat fraction (%)</b>	15.5 ± 24.0	1.22 ± 1.47	0.002	0.83
<b>CMTES (0-28)</b>	8.0 ± 5.1	0.3 ± 1.3	ns	0.23
<b>MRC lower limb</b>	95.4 ± 15.4	-0.4 ± 3.8	ns	-0.11
<b>SF-36 PF (0-100%)</b>	65.3 ± 23.2	-1.9 ± 14.9	ns	-0.12
<i>Verhamme et al, 2009,, 5 years, 46 patients</i>				
<b>Adapted CMTNS^ (0-33)</b>	11.6 ± 4.5	1.5 ± 3.0	0.003	0.49
<b>Nine-hole peg test (s)</b>	26.0 ± 18.4	2.0 ± 5.8	0.02	0.35
<b>Three-point grip (N)</b>	7 ± 32	-6.3 ± 10.5	<0.001#	-0.60
<b>Ankle dorsiflexion (N)</b>	190 ± 74	0.1 ± 28.1	ns	0.00
<b>Ankle plantarflexion (N)</b>	>250*	NA	NA	NA
<b>Ulnar CMAP (mV)</b>	4.1 ± 1.5	-0.4 ± -0.7	0.001*	-0.60
<i>Pareyson et al, 2011, 2 years, 133 patients (placebo arm)</i>				
<b>CMTNS (0-36)</b>	13.9 ± 4.3	0.5 ± 2.7	<0.05	0.19
<b>CMTES (0-28)</b>	8.6 ± 3.6	0.5 ± 2.1	<0.05	0.23
<b>CMT NCS (0-8)</b>	5.2 ± 1.6	-0.1 ± 1.6	ns	-0.06
<b>Nine-hole peg test (s)</b>	23.4 ± 5.7	0.85 ± 2.7	<0.01	0.31
<b>SF-36 PF (0-100%)</b>	62.9 ± 25.7	-1.1 ± 15.4	ns	-0.07
<b>Hand grip (N)</b>	85.8 ± 38.8	-6.9 ± 20.3	<0.001	-0.34
<b>Three-point pinch (N)</b>	65.2 ± 29.4	-3.6 ± 18.8	<0.05	-0.19
<b>Ankle dorsiflexion (N)</b>	62.8 ± 43.1	-9.8 ± 23.7	<0.001	-0.42
<b>Ankle plantarflexion (N)</b>	97.0 ± 59.7	-2.7 ± 47.7	ns	-0.06
<b>CMAP sum (mV)</b>	7.1 ± 4.1	0.2 ± 2.9	ns	0.08

Values expressed mean ± standard deviation. Non-significant (versus baseline or control) SRM depicted in grey. #: identical difference seen in controls; \*: greater reduction seen in controls ( $p=0.05$ ); +: measurement limited to 250N; ^: data for this measurement collected retrospectively; NA: not available; ns: not significant; SRM: standardised response mean, calculated from published data as mean change/standard deviation change. Standard deviation in Pareyson study calculated from published 95% confidence interval by standard statistical formulae:  $95\%CI = \text{mean} - 1.96 \text{ s.d. to mean} + 1.96 \text{ s.d.}$ ; standard error mean =  $\text{s.d.}/\sqrt{n}$ .

### **5.5.1.2 Responsiveness of outcome measures in IBM**

Progression in IBM is somewhat faster than in CMT1A, with a 10 year median interval between symptom onset and significant disability (Cortese et al., 2013). In contrast to the CMT1A group, in IBM significant change over 12 months was detected in IBM-FRS, knee extension strength, MRC scores, myometry measured knee extension and many MRI measures. Only MRI measures' SRMs exceeded 1.

In terms of longitudinal validity, quadriceps remaining muscle area correlated with change in knee extension strength (Figure 5-3). This is a demonstration of external responsiveness, longitudinal correlation between MRI-detected change and functional deficit, only demonstrated previously in Duchenne muscular dystrophy (Willcocks et al., 2014)

### **5.5.2 Baseline predictors of change**

Also of note is the longitudinal demonstration that abnormal water distribution (baseline abnormalities in STIR) leads to greater progression in fat infiltration in the subsequent 12 months. In IBM, the difference was large with 5.9% and 1.1% increase fat infiltration in muscles with marked and no baseline STIR hyperintensity respectively (Table 5-2). This difference was even apparent in CMT1A where STIR changes were not frequently seen. Post-acquisition estimation of T2-excess at baseline was also a predictor of change in FF over the subsequent 12 months. Future studies might utilise methods to separate out the fat and water T2 components at point of acquisition (Janiczek et al., 2011) so that this relationship can be investigated more directly.

Rate of change can additionally be predicted in both groups from baseline fat fraction with moderate fat fractions between 5 and 40% in IBM and 20 and 60% in CMT1A being associated with the highest rate of disease progression (Table 5-2). A similar effect has been noted in FSHD (Janssen et al., 2014). Baseline predictors of interval change may be utilised to maximise outcome measure responsiveness by selecting patients or muscles predicted to show maximum change over trial duration, whilst methods to improve reliability (Fischmann et al., 2014) can further increase responsiveness by reducing measurement-related variability.

### **5.5.3 Longitudinal insights into CMT1A**

The lack of significant change in thigh level indices in CMT1A is of no great surprise, given existing knowledge of disease distribution confirmed in the cross-sectional data in this thesis. It stands to reason that normal muscle tissue has not progressed in the one year prior to baseline, so is unlikely to progress in the one year after baseline, unless it happens to be the year where pathology reaches this level. Indeed, with the

length dependent distribution of CMT1A, one might envisage progression like water (or more accurately fat) gradually filling a pair of waders (Figure 5-6). The orange represents a zone of active denervation.

This model is largely borne out by the data presented in this thesis. The youngest patients in the study had normal fat fraction at mid-calf level, and had no progression over 12 months at this level. The oldest patients had the highest level of fat fraction, and showed little progression at this level over 12 months. A small number of muscles were STIR positive, and these muscles showed greatest progression over 12 months (2.4% vs 0.5% in muscles with normal STIR intensity). Furthermore, progression is greatest at intermediate levels of fat fraction (20-60%), peaking at 7.3% per year in the 40-60% FF range. The scatter-plot of upper calf-FF vs age (Figure 4-12) also shows a similar non-linear progression in FF when measured at a single anatomical level.

*Figure 5-6: A model of disease progression in CMT1A*



*Black represents healthy muscle, orange muscle with active denervation and white fat replaced muscle. The dotted lines show the anatomical levels analysed in this study.*

There are different ways in which this pattern of progression may be taken into consideration for a future clinical trial in CMT1A. If the sole aim was to maximise SRM, one could select patients with a similar level of severity, or indeed select the planned site for analysis based on baseline severity. The difficulty with this approach is that it may be that muscle tissue passes a point of no return, where even if innervation is re-established, on-going progression of IFA occurs due to increasing biological strain on the remaining tissue, as occurs in chronic kidney disease. On the other hand, choosing patients who are not affected at all at the anatomical level of analysis, will result in a very insensitive outcome measure. The solution may be to collect FF data over a wide field of view, from the feet to the thighs, so that disease progression can be assessed in all patients. This will however require careful consideration of how the

primary outcome measure will be defined and combined across different patients in a statistically meaningful way, whilst maintaining high responsiveness.

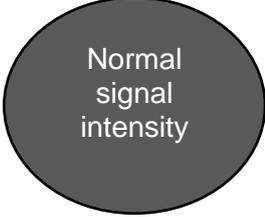
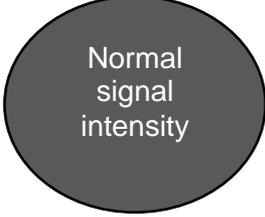
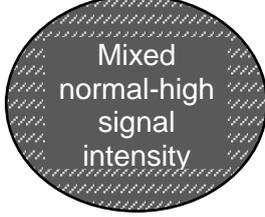
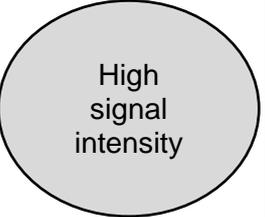
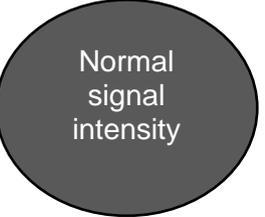
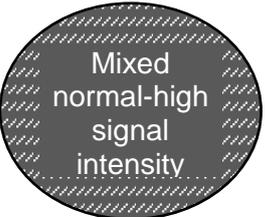
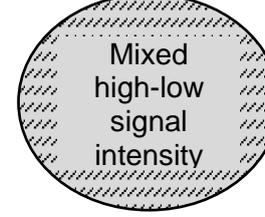
We have already added 3D-Dixon acquisitions to the ongoing scanning in the patients reported in this study and have included this in international studies in CMT1A and local studies in other hereditary neuropathies such as hereditary sensory neuropathy, with further analysis of these data ongoing. However, given that all previously tested outcome measures in CMT1A have proven entirely unresponsive to date, the demonstration of high responsiveness of muscle fat fraction at a single anatomical point in CMT1A is ground-breaking, with the clear potential to improve this even further.

#### **5.5.4 Longitudinal insights into IBM**

FSHD has more in common with IBM than may be apparent at first glance – although an inherited disease, and classified as a muscular dystrophy, prominent inflammation on biopsy may be seen on biopsy, periods of relatively rapid progression are recognised and some experts have advocated the use of immunosuppression in some cases. On the other hand, inclusion body myositis, although considered an acquired inflammatory muscle disease, defects of muscle regeneration and repair are thought to be central to the pathogenesis, and immunosuppressive or immunomodulatory interventions have been universally unsuccessful in clinical trials. Both diseases also show very specific (but different) patterns of muscle involvement, and often show asymmetry suggesting important environmental influences. FSHD may therefore be considered at the inflammatory end of the dystrophy spectrum, and IBM at the dystrophic end of the myositis spectrum.

It is interesting therefore to note the similarities between this study and the previous longitudinal quantitative MRI study in FSHD (Janssen et al., 2014). Both studies demonstrate greatest progression in fat fraction in those muscles either STIR positive or showing intermediate levels of fat-fraction at baseline. Indeed, with the overall longer duration of disease in FSHD, an hourglass distribution of FF within muscles was seen with most muscles either normal or end stage, and few with intermediate FF levels. The supposition (Figure 5-7) has been that the disease process within muscle starts with abnormal muscle water distribution, as seen in 4.4.4.2, with increased T2 and reduced MTR in muscles without significant IFA, followed (if untreated) by progressive intramuscular fat accumulation concurrent to ongoing abnormal muscle water distribution, leading ultimately to end stage fatty atrophy where there is not sufficient muscle water left to be abnormally distributed. This thesis is the first work to conclusively show that this is indeed the case.

Figure 5-7: A model of disease progression in IBM

<b>T1W</b>				
<b>STIR</b>				
<b>Stage</b>	<b>Healthy muscle</b>	<b>Early disease</b>	<b>Mid-disease</b>	<b>End-stage disease</b>

Of course, as is evident in this and previous studies, at a single time-point, different muscles within the same patient will be at different stages in this process. Indeed because of the selective involvement of muscles in IBM, this usually holds even within the same anatomical level. This has interesting implications when considering application of MRI outcome measures to clinical trials in IBM, or indeed any primary muscle disease which follows the same schema. Water sensitive measurements such as T2, ideally with separation of the confounding fat effects at acquisition stage, should be included to monitor the earliest stage of disease, and also provide an outcome measure which will potentially improve with effective therapy. Fat quantification measures, such as three-point Dixon can be used to assess for prevention of slowing of intramuscular fat accumulation, with muscles most likely to show this if STIR positive and/or with intermediate fat fraction levels (early to mid-stage disease). Healthy muscles may be most amenable to physiotherapy type interventions, whilst end-stage muscles are unlikely to be responsive to any intervention, lacking the necessary scaffolding for even stem cell therapy to be effective.

The consistent spread of muscle involvement – quadriceps before hamstrings in the thigh, medial gastrocnemius advanced even in early overall disease status in the calf – provides the opportunity to design outcome measures most likely to be responsive for a given patient group and intervention. In the ongoing follow-up of the patients in this study we have added the IDEAL-CPMG sequence (Janiczek et al., 2011), which allows

separation of fat and water effects on T2 at the acquisition stage and should provide further insights.

### **5.5.5 Longitudinal study limitations**

Outcome measure SRM derived from observational studies are only applicable to interventional studies when the intervention will affect the outcome measurement. For example, if an intervention had an effect on muscle which improved strength without an effect on muscle size or quantitative MRI variables, a functionally important benefit may be missed. Or if an intervention improved upper limb but not lower limb strength, this improvement would not be measurable with the current protocol. This relationship will need to be determined in a disease and intervention specific manner.

## **5.6 Conclusions**

MRI detected significant FF progression over 12 months in both disease groups. In the more slowly progressive CMT1A, calf-level MRI measured FF was the only measure to change significantly. In the more progressive IBM, significant changes in multiple MRI, and some clinical and myometric measures were evident, with the MRI indices being most responsive. Data from IBM patients also showed longitudinal validity, with correlation in change in remaining muscle area on MRI and myometric measured strength. Moderate baseline fat fraction and baseline abnormal water distribution predict subsequent fat fraction progression over 12 months. Quantitative MRI therefore provides valid and responsive outcome measures for future trials in CMT1A and IBM, with potential applicability for other neuromuscular diseases.

## 6 Conclusions

There has been rapid progress in the molecular understanding of many muscle-wasting neuromuscular diseases. New targets have been identified and the opportunity for testing new therapies is increasing. However, accurately and reliably monitoring disease progression is an important obstacle to successful experimental clinical trials to test new agents. We have systematically investigated the utility and reliability of quantitative MRI biomarkers to track disease progression. We have also correlated findings with important measures of patient function. The biomarkers we selected detect and quantify the common findings of water and muscle fat accumulation which both occur across a range of different muscle-wasting diseases. We studied two common exemplar neuromuscular diseases and controls.

Previous studies in NMD have quantified fat fraction (Gaeta et al., 2011; Gloor et al., 2011; Hiba et al., 2012; Willis et al., 2013; Wren et al., 2008), T2 (Gloor et al., 2011; Hiba et al., 2012; Huang et al., 1994; Kan et al., 2009; Kim et al., 2010a; Maillard et al., 2004a), MTR (McDaniel et al., 1999; Sinclair et al., 2012a) and T1 (Huang et al., 1994) in myopathic (Gaeta et al., 2011; Gloor et al., 2011; Hiba et al., 2012; Huang et al., 1994; Kan et al., 2009; Kim et al., 2010a; Maillard et al., 2004a; McDaniel et al., 1999; Willis et al., 2013; Wren et al., 2008) and neuropathic (Gaeta et al., 2011; Sinclair et al., 2012a) inherited (Gaeta et al., 2011; Gloor et al., 2011; Hiba et al., 2012; Huang et al., 1994; Kan et al., 2009; Kim et al., 2010a; McDaniel et al., 1999; Sinclair et al., 2012a; Willis et al., 2013; Wren et al., 2008) or acquired (Maillard et al., 2004a; Sinclair et al., 2012a) conditions at thigh (Gaeta et al., 2011; Gloor et al., 2011; Kan et al., 2009; Kim et al., 2010a; Maillard et al., 2004a; Willis et al., 2013) or calf (Gloor et al., 2011; Hiba et al., 2012; Huang et al., 1994; Kan et al., 2009; McDaniel et al., 1999; Sinclair et al., 2012a) level, in one (Gaeta et al., 2011; Hiba et al., 2012; Huang et al., 1994) or many (Gloor et al., 2011; Kan et al., 2009; Kim et al., 2010a; Maillard et al., 2004a; McDaniel et al., 1999; Sinclair et al., 2012a; Willis et al., 2013; Wren et al., 2008) muscles, with (Gloor et al., 2011; Hiba et al., 2012; Huang et al., 1994; Kan et al., 2009; Maillard et al., 2004a; McDaniel et al., 1999; Sinclair et al., 2012a) or without (Gaeta et al., 2011; Kim et al., 2010a; Willis et al., 2013; Wren et al., 2008) controls. The MRC Centre for Neuromuscular Diseases Prospective Cohort study is comprehensive in all these respects within a single scanning session of realistic duration.

Whilst MRI may be criticised for lack of diagnostic specificity in that acute and chronic muscle pathology have similar appearance across NMD, this is advantageous for outcome measure applications. As this study demonstrates, the same types of changes can be quantified in these two very different diseases: an acquired late-onset progressive proximal and distal myopathy and an inherited childhood-onset slowly

progressive distal predominant neuropathy. Although this results in differences in magnitude and distribution of abnormalities, their presence, direction and clinical correlations are consistent, making these outcome measures applicable across any NMD with lower limb weakness. Through selection of MRI sequences targeted to disease and intervention, MRI outcome measures can be optimised to provide maximum responsiveness for a specific clinical trial. For example, in the diseases studied, the optimal MRI protocol for CMT1A would necessarily include assessment of fat infiltration of lower limb muscles with coverage targeted to disease severity; whilst to assess if an intervention in IBM reversed acute pathological processes, additional water sensitive sequences such as T2 and MTR quantification should be included. For any NMD, understanding of basic pathomechanisms, disease distribution and treatment mechanisms will allow selection of the appropriate parts of this comprehensive protocol.

In these representative NMD, MRI provides valid outcome measure closely correlated to strength, function and disease severity. In CMT1A we have demonstrated responsiveness which far exceeds all existing outcome measures, allowing design of adequately powered clinical trials in this gradually progressive but debilitating disease. For the first time we show longitudinal evidence of correlation between decline in strength and MRI progression and also between baseline acute changes in water distribution and subsequent progression of irreversible fatty atrophy. Together the methods provide objective, non-invasive, valid, responsive outcome measures in IBM and CMT1A, with potential application to the wide spectrum of neuromuscular diseases.

## 6.1 *Specific conclusions and future directions*

### 6.1.1 Study design

For entirely novel MRI outcome measures, for example muscle diffusion imaging, it would be advised to follow the same outcome measure development process (Figure 2-1, page 38): development of technique, application to healthy controls to assess reliability, cross-sectional assessment in patients and controls to assess validity, longitudinal assessment in patients and controls to assess responsiveness, before final application in an interventional study. Each step is a necessary precursor to the next, for example without optimising reliability, responsiveness will be poor. New techniques should be run in conjunction with established methods so that their relationships, co-dependencies and relative responsiveness can be determined.

For refinements of existing MRI methods, for example 3D three-point Dixon, or more sophisticated T2 measurement methods, the same development process is required, though some aspects will be known in advance, such as distribution of abnormalities for a given disease, or expected age and gender dependencies which will influence exact trial design. Again the facility to compare established and refined method in terms of reliability, validity and responsiveness is key to demonstrating that the new method delivers on the expected improvements. Applying the same methods to new diseases may be slightly different. Qualitative imaging is sufficient to profile the disease in terms of type of pathology (acute vs chronic) and disease distribution which will inform the likely optimal protocol. On the other hand, to allow accurate sample size calculations for an interventional study, responsiveness of the protocol in that specific disease must be determined in a longitudinal study.

The process to allow the utilisation of an equivalent protocol (as far as hardware and software restraints allow) in an already studied disease at a new site, as will be necessary for phase 3 clinical trials in rare diseases, merits special consideration. More work needs to be done to demonstrate the feasibility of this; beyond the study in LGMD2I (Willis et al., 2014), data in this area are relatively lacking. In interventional studies, longitudinal change is most relevant so inevitable systematic bias in measurements between sites would be acceptable if of small magnitude. Inter-site quality control could be assessed by means of a phantom, or indeed human control or patient, scanned across multiple sites. Inter-scan reliability at each site is of greater importance and therefore needs systematic assessment. The processes here are well established in other domains, for example central nervous system imaging in multiple sclerosis trials, and need to be replicated in the peripheral nervous system domain.

The inclusion of healthy controls is vital in both cross-sectional and natural history patient studies. When healthy control data are compared across different studies, whether T2 measurement or fat quantification, it is clear that exact healthy values depend on exact scanner and sequence configuration (see 3.2.1). Therefore, in cross-sectional studies the inclusion of healthy controls is vital to establish the healthy range for the specific configuration in the study, though an equal number of controls to patients may not be necessary for this purpose. If studies are performed in children, fully age-matched controls are vital to allow correction of results for the effects of normal growth and development. This thesis provides evidence that the effect of normal ageing will not have a strong confounding effect on quantitative MRI tissue parameter outcome measures, but this should ideally be shown in longer studies, and scanning of the healthy controls in this study is continuing.

Beyond this control, data are needed in longitudinal studies to guard against systematic bias between study visits which may exaggerate or eliminate true change between time-points. The effect of the software upgrade on the T2 measurements in this study was brought to light by the unexpected and isolated change in control data between the two time points; if this had been overlooked, T2 quantification would have incorrectly appeared the most sensitive outcome measure. In randomised placebo controlled clinical trials, healthy controls are not mandatory as the placebo group provides the internal control, however inclusion of some controls for inter-site standardisation and identification of systematic bias will still increase study power. Conclusions from any study which includes no controls should be treated with caution.

## **6.1.2 MRI protocol**

The protocol presented here is the most comprehensive ever undertaken in a large longitudinal series including T1-weight and STIR qualitative imaging, and three-point-Dixon fat fraction, T1 mapping, T2 mapping and magnetisation transfer quantitative imaging, with coverage of both thighs and both calves. These methods have advanced since the study and it is always difficult to balance continuing with established methods or using the newest, more advanced but untested sequences. Whilst exact specifications will depend on disease of interest and local hardware, this thesis provides many important observations in this.

### **6.1.2.1 MRI hardware**

The first issue is field strength: 1.5 tesla as is still the mainstay of clinical practice, while a 3 tesla scanner was used in this and commonly in other research studies. There are advantages and disadvantages to both. 1.5T scanners are more widely available but 3T scanners are most commonly used in neuromuscular disease research. 3T

scanners are faster, but have significant B1 field inhomogeneity artefact, which affected the T1 mapping and MTR methods significantly in this study especially in the right thigh anteriorly, even with B1 field mapping and correction. Some methods such as three-point Dixon FF should theoretically give equivalent results at both 1.5T and 3T field strength, whereas others such as T1 recovery time are inherently dependent on field strength (Henriksen et al., 1993). The time intensive nature of many quantitative MRI sequences favours 3T MRI for research, and including scanners of different field strengths in a multi-site study would present significant but not insurmountable challenges, depending on the planned sequences.

Similar challenges exist in including scanners of the same field strength but from different manufacturers, and even different MRI models from the same manufacturer or, as this study shows, different software versions on the same scanner. Whilst these considerations were not examined in this thesis, where a single scanner was used, they are clearly an issue which needs to be addressed before large multi-site studies can be performed. This has already been investigated using three-point Dixon in LGMD2I (Willis et al., 2013) which included the MRC Centre as a site, and measuring T2 and FF across three scanners from two manufacturers in DMD (Forbes et al., 2014). What is key in any trial design is that longitudinal assessments in an individual study subject are undertaken on the same scanner so that outcome measure responsiveness is not reduced by inter-scanner differences.

#### **6.1.2.2 Scan scheduling and subject positioning**

These are aspects which were specified in this study, but not as rigorously controlled for as in some other studies. In terms of scan scheduling, MRI in patients usually occurred prior to strength testing, and if following, after at least a 1-hour resting interval, based on early studies which showed that acute T2 changes in muscle following exercise are short lived (see 3.2.2.1) and based on the good reliability of T2 measurements in healthy controls in phase one of the study, where no specific recommendations with regard to exercise were stipulated. Other studies though have shown T2 effects for up to 24 hours post exercise (Jia et al., 2015) and some protocols have had stipulations to avoid exercise for up to three days prior to scanning, though the other extreme of bed-rest also has significant effects on MRI parameters. Although no effect of exercise on T2 was found in juvenile dermatomyositis (Maillard et al., 2004a), further examination of these effects in other patient groups would provide clearer guidance as to adequate study design in this regard. Good practice until such data are available would be to perform study procedures in the same order each visit, with MRI before any strenuous physical assessments (unless that is the object of

study) and where possible ask patients to travel to assessments by the same means each visit.

Subject positioning in this study was only specified as far as “feet-first supine” but one of the qualitative observations through the study was that on occasion muscle deformation due to different calf positioning on the table, or rotation of the lower limb about the long axis, significantly affected the axial outline of muscles on baseline and follow up scans. As long as tissue deformation occurs within the same axial plane, the same volume of tissue will be assessed, though even then defining equivalent ROI between two scans is made much more difficult with distorted anatomy. For this reason we undertook a positioning study looking at this question specifically (Fischmann et al., 2014) and found that relaxed or more supported limb positioning methods were equally reproducible, but that reproducibility reduced significantly if different positioning methods were used. In a clinical trial it could not be guaranteed that the same radiographer performs baseline and follow-up scans for each patient, so some specification to subject positioning is warranted.

### **6.1.2.3 Anatomical coverage and slice positioning**

Whilst whole-body muscle MRI is gaining an increasing following in neuromuscular MRI for diagnostic purposes (Ohana et al., 2014; Quijano-Roy et al., 2012), lower limb MRI remains the mainstay of publications describing pattern of muscle involvement in different diseases (Wattjes et al., 2010) and is certainly the most common area imaged using quantitative studies (Gaeta et al., 2011; Gloor et al., 2011; Huang et al., 1994; Kan et al., 2009; Kim et al., 2010a; Maillard et al., 2004a; McDaniel et al., 1999; Willis et al., 2014; Wren et al., 2008). The MRI trade-off between coverage and signal-to-noise ratio for a fixed scanning time, together with additional technical difficulties of scanning other body parts, such as trying to position the upper limb at the centre of the bore where the magnetic field is most homogenous, or the additional breathing and cardiac motion artefacts in the trunk, are likely to result in lower limb protocols remaining most popular for quantitative MRI except in specific circumstances. For example recent quantitative studies in non-ambulant boys with Duchenne utilised an upper limbs protocol (Hogrel et al., 2016; Ricotti et al., 2014). The choice between pelvis, thigh, calf and foot muscle will depend upon the disease being studied and the disease severity. For example, in IBM just thigh imaging would be potentially adequate if an intervention is expected to affect all muscles equally, whilst in CMT1A foot, calf or thigh imaging would be most appropriate for early, moderate or late stage disease respectively. The pattern of involvement can be adequately examined based on qualitative MRI data published in the literature, or when unavailable preformed as an initial pilot study.

Slice positioning is absolutely critical to test-retest reliability in patients due to the inhomogeneous distribution of pathology within a muscle, for example proximal-distal gradients within a muscle in CMT1A (Gallardo et al., 2006). We recently presented data from this study showing proximal-distal gradients of FF within calf muscles in CMT1A patients of up to 8%/cm (Evans et al., 2015), which has clear implications regarding the reliability of FF as an outcome measure if near-identical axial slice positioning is not achieved between scans. In this thesis, block positioning during acquisition was based on surface anatomical landmarks, which were inconsistently assessed and measured, resulting in a maximum inter-scan difference in slice positioning of 10cm from post-hoc assessment of bony landmarks, and a mean inter-scan difference in slice positioning of  $0.4 \pm 3.2$ cm, with a mean absolute inter-scan difference of  $2.3 \pm 2.2$ cm. In this thesis, it was possible to adjust for this shifted imaging block during the analysis phase due to the total block size of at least 19cm on all sequences (10 slices, 1cm slice thickness, 1cm slice gap). ROI were drawn on the repeat slice closest anatomically to the slice used for the baseline ROI, based on measurement from an imaged bony landmark, the lateral tibial plateau (Figure 6-1).

For more advanced sequences such as IDEAL CPMG (Janiczek et al., 2011) where only a small number of slices is possible due to long acquisition time, accurate block positioning is vital to ensure there is overlap between the volume of tissue imaged at baseline and follow up. With the identification of this issue, we showed that reliability of anatomical localisation was much improved when clock location was based on a fixed distance from a bony landmark such as the knee joint (Fischmann et al., 2014). This is therefore the method advocated in future studies.

The observed gradients also raise issues concerning inter-slice distance. In this study there was a 2cm total distance between adjacent axial slices (1cm slice thickness, 1cm gap), which meant that even with careful post-acquisition assessment of slice localisation in the proximal-distal direction, up to 1cm difference in sliced localisation is unavoidable. This difference occurs when slices are perfectly interleaved between baseline and follow-up. Whilst increasing number of slices analysed can help if the inhomogeneity of intramuscular fat accumulation is random, with a gradient, increasing the number of slices analysed does not improve the accuracy (see Figure 6-2). A potential solution is to image and analyse on the entire volume of muscle tissue, but this is very labour intensive, boundary definition is more difficult at muscle ends, and different sampling within that volume will still affect results. A method to minimise this is to reduce inter-slice distance, and we are now investigating the utility of a “3D” three-point Dixon sequence with no slice gap and 5mm slice thickness in the proximal-distal axis, which means that slices within 2.5mm of each other can be assessed

longitudinally. An analysis of the effects of these considerations on test-retest reliability is further discussed in 6.1.3.2 and depicted in Figure 6-1 and Figure 6-2.

#### **6.1.2.4 MRI sequences**

Two sequences in this study were included primarily to measure intramuscular fat accumulation: three-point Dixon and T1 mapping. T1 mapping performed worse on almost every metric: presence of artefacts or unusable maps, lower test-retest reliability in volunteers, evidence of systematic bias on longitudinal study and reduced responsiveness on longitudinal analysis. Compared with a sparse literature on T1 mapping, there is a large and ever increasing body of literature utilising three-point Dixon, or similar techniques based on the same principle, which show consistent reliability, validity and responsiveness of fat quantification methods across a range of neuromuscular diseases, including in multi-site studies. Without doubt therefore, chemical shift based fat-water separation quantification methods such as three-point Dixon are favoured. Without further evidence, T1 mapping would not be recommended in future clinical trials.

There are many different chemical shift based fat-water separation quantification methods, and even within a single method such as three-point Dixon, further refinements are possible, such as 3D coverage or accelerated acquisition algorithms (Hollingsworth, 2015). A major factor in test-retest reliability in patients is having a small distance between axial slices (see 6.1.3.2). Only one previous paper has compared methods directly in cross-sectional analysis, and found two-point Dixon to be superior to a steady state free precession method (Gloor et al., 2011). No studies have compared methods longitudinally. The exact choice of sequence will therefore depend on MRI hardware and software platforms available for the study.

The measures sensitive to abnormal water distribution (acute pathology) in this study were T2 mapping and magnetisation transfer imaging. Overall these performed similarly: as applied both were confounded by fat effects, MTR had overall more artefacts, T2 was affected by a routine software upgrade, both showed changes in early disease in muscle without significant fat accumulation. For both there are more advanced methods to separate the confounding effect of fat (Janiczek et al., 2011; Sinclair et al., 2010), though these are significantly longer for the same anatomical coverage. In early studies where full characterisation of changes in muscle tissue parameters are needed, it would still be reasonable to include both, but in clinical trials where scan duration is of greater importance and applicability across multiple scanners is important, just a T2 mapping measurement might be preferred. More advanced T2 sequences such as an increased number of echo times, or utilising methods to

suppress or separate fat signal at the point of acquisition, would be advised in future studies.

T1-weighted sequences and STIR or equivalent sequences remain the mainstay of diagnostic imaging. Their inclusion in quantitative studies remains helpful to provide a cross-reference for quantitative assessments of IFA and abnormal muscle water, as an internal control of the quantitative sequences validity (convergent validity). The converse question could also be proposed: whether quantitative sequences could be used for diagnostic purposes – for example a three-point Dixon type sequence to demonstrate pattern of muscle involvement, or a T2 water mapping sequence to measure disease activity, to guide treatment or select sites for biopsy in inflammatory myopathies. There is no theoretical reason why this could not be the case, but there remain significant practical obstacles to implementation.

Whether both three-point Dixon (or equivalent) and T2 mapping methods both need to be included in a study protocol naturally depends on the study purpose and the disease being studied. However, even in CMT1A, one of the most gradually progressive neuromuscular diseases, evidence of active disease was identified on T2 sequences, so the inclusion of both types of sequences would provide complementary information and would be recommended.

#### **6.1.2.5 Novel methods which might be considered for future studies**

In addition to the MRI methods used in this study and refinements thereof, there are novel approaches which are worthy of consideration in future studies, specifically diffusion tensor MRI and MR spectroscopy.

##### **6.1.2.5.1 Diffusion tensor MRI**

Diffusion tensor imaging is useful for studying organ structure and is widely used in imaging of the central nervous system and brain (Le Bihan et al., 1986). Diffusion tensor imaging is based on the restriction of the diffusion of water by cell membranes and other structures. This restriction results in an apparent diffusion co-efficient lower than the free diffusion coefficient. Diffusivity is orientation dependent in elongated structures. Thus in skeletal muscle diffusivity would be expected to be greater longitudinal to the axis of the myocytes (Damon et al., 2016; Longwei, 2012).

Diffusion tensor MR has been applied to skeletal muscle in healthy volunteers including calf (Saupe et al., 2008; Sinha et al., 2006; Steidle and Schick, 2015) and thigh (Budzik et al., 2007) musculature. It has further been examined in looking at the effects of ageing (Sinha et al., 2015) and exercise (Cermak et al., 2012) in humans, and denervation in a rat model (Ha et al., 2015). Thus, diffusion tensor MRI parameters

may be sensitive to early changes in myocyte structure and integrity and is worthy of further study in neuromuscular diseases.

#### **6.1.2.5.2 MR Spectroscopy**

As already discussed, H1-spectroscopy can be used to quantify fat and has been applied to neuromuscular diseases (see 3.2.1.3.1). H1-spectroscopic methods also allow the calculation of the T2 time of the water component independent of the fat component in DMD (Forbes et al., 2014). However, employing spectroscopic MR to other elements such as phosphorus and sodium, provide additional means of assessing skeletal muscle function.

Sodium plays an important role in cellular integrity and function, with a concentration gradient across the cellular membrane maintained by a sodium-potassium pump (Constantinides et al., 2000). Sodium spectroscopy has been used to estimate the total sodium concentration in normal calf muscle (28.4 mmol/kg) and has been shown to be elevated post exercise and in dystrophic muscle (Constantinides et al., 2000). A particularly enticing application of sodium spectroscopy is the skeletal muscle channelopathies, inherited muscle diseases where the primary genetic defects are in the ion channels in the muscle cell membrane, including mutations in the voltage gated sodium channel (SCN4A) which result in a periodic paralysis phenotype (Raja Rayan and Hanna, 2010). Indeed sodium spectroscopy has been employed by Jurkat-Rott and colleagues in patients with hypokalaemic periodic paralysis showing increased sodium concentration within calf muscles, which normalised following treatment with acetazolamide (Jurkat-Rott et al., 2009). Sodium spectroscopy may therefore provide novel insights and treatment responsive biomarkers in neuromuscular diseases and warrants further investigation.

### **6.1.3 Analysis methods**

The key to optimising reliability of analysis is to ensure the same volume of tissue is selected at baseline and follow up, as any difference will result in variability which will reduce study responsiveness. Ensuring similar patient positioning is key at the acquisition stage (see 6.1.2.1). During ROI definition, capturing the same tissue needs to be considered in both the proximal-distal direction (z axis) and axial plane (xy axes).

#### **6.1.3.1 General considerations**

As in this study, ideally all ROI should be drawn by a single observer. As has been demonstrated in studies assessing reliability, intra-observer reliability is greater than inter-observer reliability (see 3.2.4), and regardless, inter-observer variation can be removed entirely by only using one observer. In a large study where that is not

possible, one observer should draw all ROI for a given subject to eliminate inter-observer variation from the longitudinal assessments in each subject.

The observer should be appropriately experienced or trained in neuromuscular MRI, although there are no studies specifically addressing this question. The observer should be blinded as to the disease status of the subject, and all other clinical details, to eliminate any potential bias. This is another benefit of including a control group – the observer will remain blinded to healthy control versus minimal disease such as the mildly affected CMT1A subjects in this thesis. The observer should be blinded to visit number of the scan to eliminate any potential bias when drawing repeat ROI. Exact positioning of ROI in heterogeneous tissue or close to muscle boundaries can have a significant effect on fat fraction, and knowledge of the scan order could lead to subconscious bias, which can easily be eliminated by blinding this information. As the ROI in our study were performed for phase 2 then for phase 3, blinding to visit number was not done for this thesis, which is a weakness. Conversely ROI should be drawn with reference to other ROI drawn in the same subject to ensure equivalent ROI are drawn between scans. Some inter-muscle boundaries are not anatomical, for example between the three vasti muscles, and drawing ROI with reference to each other allows consistency in inter-muscle definition, or in the case of small ROI, equivalent placement within the muscle.

In this study ROI were drawn on one of the raw Dixon acquisitions (se=345ms) rather than generated maps. This sequence has good anatomical detail but relatively less fat-muscle contrast, which aids ROI definition, particularly in patients with severe disease. On fat fraction maps for example, both muscle fascia and healthy muscle tissue have near zero fat fraction, so these images are not easy to use to define inter-muscular boundaries, particularly in skinny patients. Software should allow ROI to be saved so that a data checking and ROI correction process is possible, as in this study.

#### **6.1.3.2 Z-axis considerations**

The present of disease inhomogeneity within a muscle, in particular a gradient within the muscle, makes consistent localisation in the z-axis crucial. As above (6.1.2.3), this should be applied at scanner set-up phase, most reliably by scanning at a fixed distance from a bony landmark – in lower limbs most easily the knee joint, or as specifically used in this study, the right lateral tibial plateau. The same process can be applied at the analysis phase by defining the location of the anatomical landmark on the scout imaging, then calculating the slice closest to the desired distance proximal or distal to the knee joint. This should be calculated for the right and left limbs separately as any degree of change in pelvic tilt between scans will affect the relative positions of the right and left leg, and ROI may need to be drawn on a different slice number

between right and left legs. There are different considerations in children where the muscle tissue is a moving target due to participant growth in between scans. In children therefore a fixed proportion of the distance between identified bony landmarks may be more appropriate.

The effect of different slice positioning and analysis methods on test-retest reliability is considered in a hypothetical patient muscle with a fat fraction gradient of 4%/cm, the median gradient seen in an analysis of the CMT1A patients in this study (Evans et al., 2015), and is depicted graphically in Figure 6-1 and Figure 6-2. Using the same slice within the block as acquired (5th of 10 slices) for analysis would have the greatest effect on the measured value. Figure 6-1-A depicts the effect with the median inter-scan block acquisition displacement of 3cm seen in this thesis. This results in a 12% difference on repeat imaging of the same muscle. This degree of error would greatly limit the ability of MRI to detect small changes in FF longitudinally.

The method use in this thesis was to define ROI on the anatomically closest slice. However, with 2cm between slices, the maximum distance between analysed slices would still be 1cm (Figure 6-1-B), resulting in a 2% test-retest difference.

Figure 6-1: Effect of slice selection on test-retest reliability

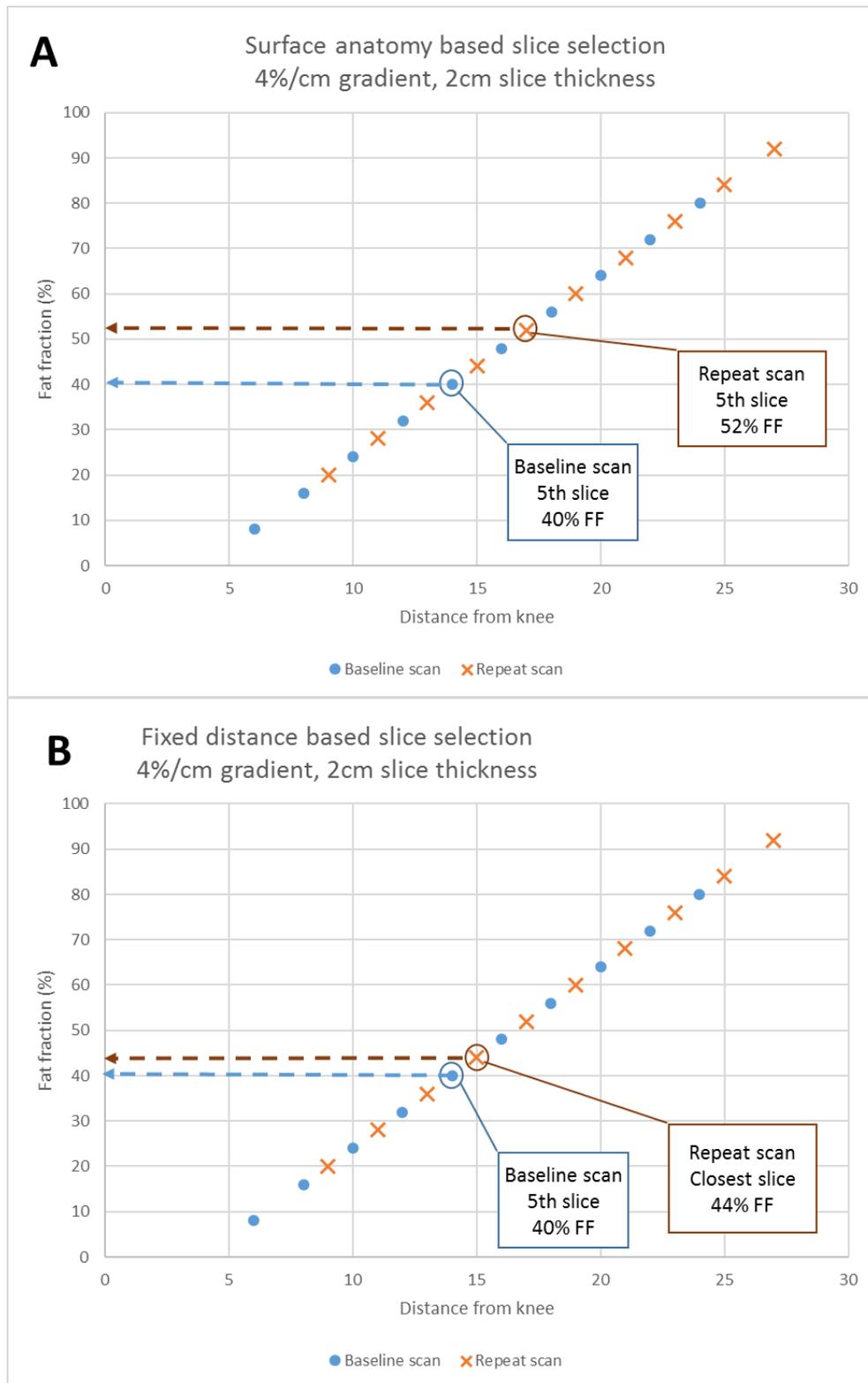
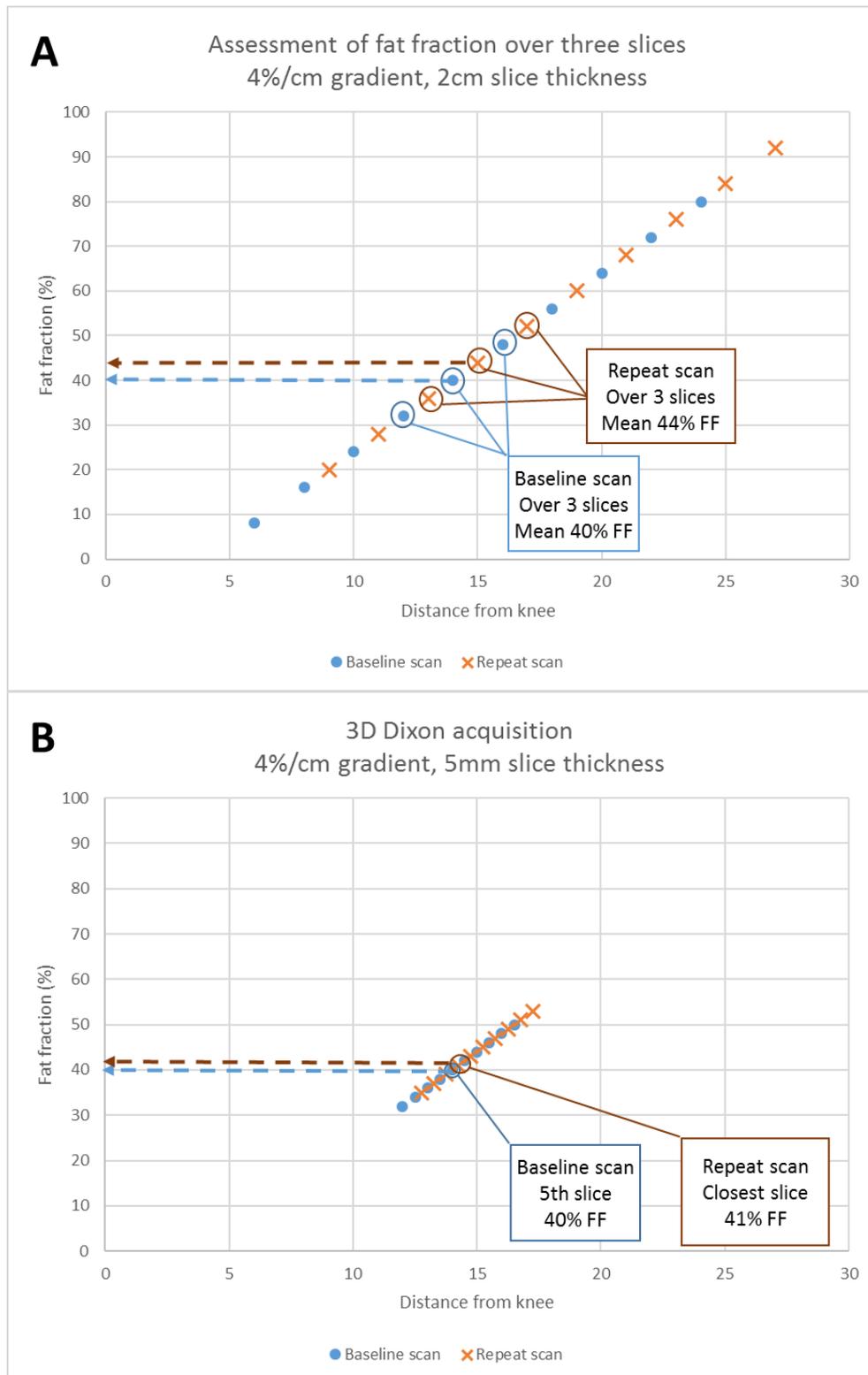


Figure 6-2: Methods to Improve test-retest reliability



One method to try and improve reliability would be to increase the number of slices analysed, for example to analyse three slices rather than a single slice. Depending on the relationship between slice position and disease distribution this may in fact worsen test-retest reliability, and in our hypothetical patient with a constant gradient, there is no change in the test-retest difference of 2% (Figure 6-2-A). Rather the best method to improve reliability is to reduce slice thickness, so that the z-axis displacement is minimalised. In our model, reducing the slice thickness to 5mm, reduced the inter-scan difference to 0.5% (Figure 6-2-B). This is indeed the method we have applied in ongoing scanning of this patient cohort, as well as in an international study extension, and would be recommended in any future studies where there is disease inhomogeneity within muscle tissue.

### **6.1.3.3 X-Y plane considerations**

Different studies have utilised different sized regions of interest within the x-y plane: broadly small, subtotal and whole. The best to utilise will depend on the aim of analysis. Where correlation with strength is desired, whole ROI should be used to allow inclusion of cross-sectional muscle area, a major determinant of muscle strength. A disadvantage of using whole ROI is the partial volume effect, where voxels which are part muscle and part fascia/subcutaneous fat are included which can skew the results. In this study a systematic bias was present between observers for whole ROI muscle fat fraction in volunteers, with FF from one observer ~1% higher, which when assessed was due to inclusion of more voxels within the ROI at the muscle boundary. Conversely when small ROI are used, even with specific details of where within the muscle the small ROI should be drawn, small changes in the size or position will affect the result, especially in patients with heterogeneous disease within a muscle.

In the T1, T2 and MTR methods used in this study, there were sometimes artefacts at tissue boundaries which limited usability of these ROI for these sequences. Furthermore, in order that the same volume of tissue could be compared across sequences, a single set of ROI was drawn on one of the Dixon acquisitions, which were then transferred to co-registered maps of the other sequences. All these were manually checked to ensure they lay within the appropriate muscles. If whole muscle ROI were used, any movement between sequences would have affected the extracted muscle values significantly. The main advantage of small ROI is that an area of tissue without vessels or intramuscular septae or tendons can be chosen. An intermediate option is a sub-total ROI which is drawn a set number of voxels inside the perceived tissue boundary. Finally, in the CMT1A patients, whole FF was more responsive than small FF due to less variability of change – presumably due to greater reliability of

whole muscle FF. For future studies, size of ROI will depend on the planned purpose of the extracted data, with whole ROI preferred for FF outcome measures.

#### **6.1.3.4 Data checking**

Robust data checking procedures were undertaken in this study and would be recommended for future studies:

- All ROI were checked directly on parameter maps to ensure accuracy.
- All parameter maps were visually checked for artefact, and ROI affected by artefact excluded from all analysis.
- Visual assessment was performed of scatter plots between quantitative parameters (e.g. T2 vs FF) for outliers whose corresponding maps/ROIs were re-reviewed for artefact and ROI placement.
- Visual assessment was performed of scatter plots of baseline versus follow up values for outliers whose corresponding maps/ROIs were re-reviewed for artefact and ROI placement.
- Examination of control data for possible systematic bias between scans (eg the T2 upgrade effect observed in this study).

All these procedures should take place whilst still blinded to subject group status and scan order to avoid the potential for observer bias.

A transparent data pipeline also needs to be in place: from saving ROI files so they can be re-examined, clear recording of data points excluded due to artefact, automated data extraction and collation processes to eliminate transcription errors, checking of a proportion of final data back to results directly derived from source data to ensure fidelity of data pipeline, keeping the final raw data set separate from any data analysis processes so its integrity is not compromised, and full recording of statistical analysis performed (in this case data files and output files from SPSS).

#### **6.1.4 Outcome measure selection and trial design**

In future clinical trials diseases, what specific MRI measure should be selected as the primary outcome, or one of the secondary outcome measures? For a natural history study, exploratory analysis to determine the best outcome measure is reasonable (though still should be specified in advance), but for a clinical trial to be statistically valid any planned primary or secondary outcome measure should be defined before the study commences.

##### **6.1.4.1 CMT1A**

To consider the specific diseases studied first, in CMT1A these data demonstrate that the only sequence which showed significant change was the three-point Dixon,

specifically whole muscle ROI across all muscles from a single axial slice at calf level. Whilst this measure would therefore be the clear one to use in CMT1A based on the data in this thesis, there are many ways the responsiveness of this measurement might be improved. We can consider the formula for standardised response mean to see how this might be achieved, either by increasing the mean change or reducing the standard deviation of the change.

$$\text{SRM} = \text{mean change} / \text{standard deviation change}$$

This thesis has shown methods to increase mean change on fat fraction. Change in fat fraction is greatest in muscles with intermediate fat fraction, therefore by selecting patients or muscles with intermediate fat fraction at baseline, the greatest degree of change over 12 months would be expected. The options here would be to enrol patients of a similar disease severity, with intermediate fat fraction at mid-calf level, either measured directly or predicted based on clinical assessment or qualitative imaging. The alternative is to include a broader spectrum of patients but to use fat fraction at an anatomical level appropriate to the patients as the outcome measure: foot for mildly affected, calf for moderately affected and thigh for severely affected. Careful statistical consideration would be required as to how to combine measures at different anatomical levels. One advantage of muscle FF in this regard over muscle strength or neurophysiology variables is that the normal range (<5%) and disease spectrum (5-100%) are similar in skeletal muscle regardless of calf or thigh location (see 3.2.1.3 and 3.4.3). Muscles with STIR positivity in addition to moderate fat fraction have the greatest change, but these are relatively infrequent in CMT1A so would limit recruitment; however it would be interesting to focus on the response to treatment in these muscles. The final means to increase mean change is to increase duration of follow-up, and scanning our cohort of CMT1A patients over five years is ongoing.

There are two methods of reducing standard deviation of change. One is to reduce apparent change due to poor reliability. Optimising all aspects of the protocol as outlined in the above sections is important in this regard, most crucially in CMT1A by ensuring accuracy of ROI placement in the z-axis (distal-proximal direction) due to the presence of a gradient, by utilising the smallest slice thickness practical. Beyond this, standard deviation of change may also be reduced by selection of a patient population and outcome measure where a consistent change in the outcome measure is expected. For example, a patient population with a mean change over 12 months of 1% FF, with a standard deviation of change of 0.5%, has an SRM of 2; this is more responsive than a patient population with a mean change over 12 months of 4% and a standard deviation of change of 4% with an SRM of 1. Therefore, purely maximising degree of change may not be optimal if at the expense of a wide degree of variability.

On the other hand, including very mild patients who show no progression, has a doubly negative effect on the responsiveness by reducing the mean change AND increasing the variability of change.

The final consideration is that a very responsive outcome measure based on SRM assessment is of no use for study power if the intervention has no effect on that measure:

$$N=16/(SRM \times E)^2$$

Whilst E may be 0 because the intervention is ineffective on the disease process as a whole, it may also be low if the intervention has no effect on the disease stage being measured. One potential concern in maximising SRM as outlined above is that muscles with 20-60% FF may have passed a pathological point of no return where even halting any further progression of denervation does not prevent further fatty atrophy. Indeed, there may be a point in the pathogenesis of fatty atrophy where even restoration of normal nerve supply, would not prevent further progression to end stage muscle. The difficulty is that as there is currently no effective treatment in CMT1A these thresholds cannot be directly investigated. The advantage of MRI is that anatomical coverage can include the entire lower limbs or indeed the entire body (Horvath et al., 2015), and thus different disease stages can be examined in every individual. Whilst in clinical trials the primary outcome measure needs to be pre-specified, treatment response at different pathological severities could be included as secondary outcome measures, to identify any treatment effect present.

Putting all these considerations together, the best balance between outcome measure responsiveness and disease responsiveness to treatment might be at the point of early but quantifiable pathology, roughly 5-20% fat fraction. Baseline scans could be used to determine where in the lower limbs (thighs/calves/feet) fat fraction was in this range, and change in fat fraction at this level could be used as the primary outcome measure, with secondary outcome measures including assessment of fat fraction at other anatomical levels. If feasible, T2 water measurement as a marker of early and active disease should also be included as a secondary outcome measure.

#### **6.1.4.2 IBM**

In IBM, there were both MRI and clinical measures which changed significantly over 12 months. However, no myometry measure changed in patients both compared with baseline and compared with controls (see Table 5-1), which highlights some of the difficulties in using myometry as an outcome measure. The IBM-FRS is undoubtedly an important outcome measure to include in future clinical trials in IBM, but whilst some improvements might be possible, for example utilising Rasch analysis to convert to an

interval scale, however with MRI many refinements are possible to improve responsiveness even further.

As in CMT1A, SRM may be increased by increasing mean change or reducing standard deviation of change. In particular the same considerations with regard to optimising reliability as outlined in the sections above apply. IBM however has three advantages over CMT1A in utilisation of MRI outcome measures. First, disease progression is faster in IBM, both clinically and on MRI, with progression in fat fraction more than double that seen in CMT. Second, a greater spectrum of disease stages is usually seen at a single anatomical level, as IBM displays a non-length dependent disease pattern with selective and sequential muscle involvement at a single level. For this reason, mean fat fraction at a single level such as mid-thigh may detect change across a wide range of disease severity, with progression in quadriceps fat fraction in early disease and progression in hamstrings at a later disease phase. Third, there is a greater degree of acute pathology (abnormal muscle water) to quantify which would potentially be reversible and is predictive of subsequent disease progression, which would allow baseline stratification.

In a clinical trial therefore, a detailed yet short protocol could be undertaken at a single level with high resolution quantification of intramuscular fat accumulation as the primary outcome measure, and sophisticated quantification of the T2 water component a secondary outcome measure to assess reversibility of acute changes with treatment.

### **6.1.5 Application to clinical trials of other diseases**

Whilst this thesis cannot determine the optimal means to utilise quantitative MRI in all neuromuscular diseases, by including such different diseases as CMT1A and IBM, it does suggest that these methods will have wide applicability across most neuromuscular diseases where progressive muscle weakness is a feature. This is very much supported by a rapidly increasing body of literature with both cross-sectional (4.2.2) and longitudinal (5.2) data, using similar methods across a wide range of neuromuscular diseases. General lessons learnt from this study would apply to any clinical trials in other neuromuscular diseases:

1. The need for a healthy control group in both cross-sectional and natural history studies, and probably clinical trials.
2. The importance of a comprehensive protocol measuring both acute water abnormalities and chronic intramuscular fat accumulation in cross-sectional and natural history studies to allow full characterisation of quantifiable disease processes and therefore the best outcome measures for clinical trials.

3. The importance of optimising reliability as far as possible, in particular with regard to ensuring the volume of tissue being analysed across time points is the same in both the x-y plane and in the z-axis.

The work from this thesis informed the selection and refinement of sequences utilised in quantitative muscle MRI protocols for subsequent studies at the MRC Centre for Neuromuscular Diseases. These are examples of how a quantitative MRI protocol can be tailored to specific disease and trial design factors.

#### **6.1.5.1 Quantitative MRI protocol in hypokalaemic periodic paralysis**

The skeletal muscle channelopathies provide two attractive characteristics for studying quantitative muscle MRI: transcellular fluid shifts are a direct consequence of the underlying genetic defects, and effective treatment of these fluid shifts exists (Jurkat-Rott et al., 2009; Raja Rayan and Hanna, 2010). The protocol designed therefore included sequences to measure muscle water parameters as accurately as possible. Sequences performed at mid-thigh and mid calf-level were: standard T1-weighted sequence; semi quantitative STIR sequence normalized to saline reference (Jurkat-Rott et al., 2009); 3-point Dixon fat quantification; IDEAL-CPMG (Janiczek et al., 2011) with independent quantification of muscle water T2; and MTR imaging. Other modifications of the protocol based on the results from this thesis included fixed-distance from bony landmark slice positioning, and reduced interslice distance to 10mm. The rapid response to treatment demonstrated in this condition (Jurkat-Rott et al., 2009), meant that a 4 week interscan interval was employed.

#### **6.1.5.2 Quantitative MRI protocol in Hereditary Sensory Neuropathy**

Hereditary sensory neuropathy type 1 is caused by mutations in the SPTLC1 (Serine Palmitoyltransferase Long Chain Base Subunit 1) gene. Although the initial feature is of sensory loss, length dependent motor weakness is also common as the disease progresses, and the condition can be thought of as on the CMT spectrum (Reilly et al., 2011). The pathogenesis of the condition suggests oral serine supplementation as an effective therapy, and has been utilized in both mouse models and in humans (Garofalo et al., 2011). Unfortunately it is a very rare disease which severely limits the potential size of any clinical trial, necessitating outcome measures optimized for maximum responsiveness.

We therefore included quantitative lower limb muscle MRI as an outcome measure in a natural history study with a 12 month inter-assessment interval. Due to the similarity with CMT, fat quantification was selected as the primary outcome measure, but the three point-Dixon sequence employed in this thesis was further refined by employing full 3D coverage with 5mm slice thickness in the z-axis; and anatomical coverage of the

foot, whole lower leg and distal thigh to mitigate against the floor effect noted in this thesis. The protocol also included standard T1w and STIR sequences to allow assessment of convergent validity. Preliminary analysis of the longitudinal data showed that MRI provided the only responsive outcome measure in this disease (Evans et al, in preparation).

### **6.1.5.3 Quantitative muscle MRI protocol in other neuromuscular diseases**

As demonstrated by the two examples above, the challenges for each specific disease will be different. Another example is whilst in IBM there is concern regarding the confounding effects of healthy ageing, in Duchenne the concern is over the confounding effects of normal growth. The solutions are often similar – in this case inclusion of an appropriate control group. The exact role of MRI within the drug development pipeline would also vary by disease. Often it may be used in phase 2 clinical trials where it can provide sensitive measures of efficacy and valuable insights into the nature and distribution of treatment response, whereas in many diseases it may be chosen as a secondary outcome measure in pivotal phase three trials. On the other hand, for slowly progressive conditions such as CMT1A, where no other responsive outcome measure exists, an argument can be made that MRI measured intramuscular fat accumulation should be used as the primary outcome measure in phase three trials, as other current outcome measures would be futile. This is a stance which would require careful dialogue with the drug licensing agencies such as the FDA and European Medicines Agency.

## **6.2 Final conclusions**

We are at the dawn of an exciting new era of treatment in neuromuscular diseases with medicines which alter the natural history of the disease being licensed, such as enzyme replacement therapy in Pompe disease (van der Ploeg et al., 2010) and ataluren in Duchenne muscular dystrophy (Bushby et al., 2014), with other gene therapies in late stage clinical trials. Pivotal to these studies will be valid, reliable and responsive outcome measures to allow accurate identification of effective treatments and to exclude ineffective treatments.

This thesis has demonstrated the enormous potential of quantitative MRI to provide such outcome measures in two exemplar neuromuscular diseases: inclusion body myositis and Charcot-Marie-Tooth disease type 1A. Three-point Dixon fat-quantification provides a robust measure of chronic muscle pathology whilst T2 and magnetisation transfer methods probe acute water changes within skeletal muscle. Test-retest reliability is excellent and can be improved further; construct, criterion and

longitudinal validity have been demonstrated; and responsiveness exceeds all other outcome measures assessed. Quantitative MRI should be considered for inclusion in all natural history studies and clinical trials in neuromuscular diseases.

## 7 Publications related to this PhD

**Morrow, J.M.**, Sinclair, C.D.J., Fischmann, A., Reilly, M.M., Hanna, M.G., Yousry, T.A., and Thornton, J.S. (2014). Reproducibility, and age, body-weight and gender dependency of candidate skeletal muscle MRI outcome measures in healthy volunteers. *Eur. Radiol.* 24, 1610–1620.

**Morrow, J.M.**, Sinclair, C.D.J., Fischmann, A., Machado, P.M., Reilly, M.M., Yousry, T.A., Thornton, J.S., and Hanna, M.G. (2015). MRI biomarker assessment of neuromuscular disease progression: a prospective observational cohort study. *Lancet Neurol.*

**Morrow, J.M.**, and Reilly, M.M. (2015). Early detection of nerve injury in transthyretin-related familial amyloid polyneuropathy. *Brain J. Neurol.* 138, 507–509.

**Morrow, J.M.**, Matthews, E., Raja Rayan, D.L., Fischmann, A., Sinclair, C.D.J., Reilly, M.M., Thornton, J.S., Hanna, M.G., and Yousry, T.A. (2013a). Muscle MRI reveals distinct abnormalities in genetically proven non-dystrophic myotonias. *Neuromuscul. Disord. NMD* 23, 637–646.

**Morrow, J.M.**, Reilly, M.M., and Hanna, M.G. (2013b). Reliability and accuracy of skeletal muscle imaging in limb-girdle muscular dystrophies. *Neurology* 80, 2276.

**Morrow, J.M.**, Pitceathly, R.D.S., Quinlivan, R.M., and Yousry, T.A. (2013c). Muscle MRI in Bethlem myopathy. *Case Rep.* 2013, bcr2013008596–bcr2013008596.

Cottenie, E., Menezes, M.P., Rossor, A.M., **Morrow, J.M.**, Yousry, T.A., Dick, D.J., Anderson, J.R., Jaunmuktane, Z., Brandner, S., Blake, J.C., et al. (2013). Rapidly progressive asymmetrical weakness in Charcot-Marie-Tooth disease type 4J resembles chronic inflammatory demyelinating polyneuropathy. *Neuromuscul. Disord. NMD* 23, 399–403.

Finlayson, S., **Morrow, J.M.**, Rodriguez Cruz, P.M., Sinclair, C.D.J., Fischmann, A., Thornton, J.S., Knight, S., Norbury, R., White, M., Al-Hajjar, M., et al. (2016). Muscle MRI in congenital myasthenic syndromes. *Muscle Nerve.*

Fischmann, A., **Morrow, J.M.**, Sinclair, C.D.J., Reilly, M.M., Hanna, M.G., Yousry, T., and Thornton, J.S. (2014). Improved anatomical reproducibility in quantitative lower-limb muscle MRI. *J. Magn. Reson. Imaging JMRI* 39, 1033–1038.

Pitceathly, R.D.S., Tomlinson, S.E., Hargreaves, I., Bhardwaj, N., Holton, J.L., **Morrow, J.M.**, Evans, J., Smith, C., Fratter, C., Woodward, C.E., et al. (2013). Distal myopathy with cachexia: an unrecognised phenotype caused by dominantly-inherited mitochondrial polymerase  $\gamma$  mutations. *J. Neurol. Neurosurg. Psychiatry* 84, 107–110.

Pitceathly, R.D.S., **Morrow, J.M.**, Sinclair, C.D.J., Woodward, C., Sweeney, M.G., Rahman, S., Plant, G.T., Ali, N., Bremner, F., Davagnanam, I., et al. (2016). Extra-ocular muscle MRI in genetically-defined mitochondrial disease. *Eur. Radiol.* 26, 130–137.

Ricotti, V., Evans, R.B., Sinclair, C.D.J., **Morrow, J.M.**, Butler, J.W., Janiczek, R.L., Hanna, M.G., Matthews, P.M., Yousry, T.A., Muntoni, F., et al. (2014). Upper limb muscle MRI fat-water quantification and clinical functional correlation in non-ambulant Duchenne muscular dystrophy. *Neuromuscul. Disord.* 24, 839–839.

Sinclair, C.D.J., Miranda, M.A., Cowley, P., **Morrow, J.M.**, Davagnanam, I., Mehta, H., Hanna, M.G., Koltzenburg, M., Reilly, M.M., Yousry, T.A., et al. (2011). MRI shows increased sciatic nerve cross sectional area in inherited and inflammatory neuropathies. *J. Neurol. Neurosurg. Psychiatry* 82, 1283–1286.

Sinclair, C.D.J., **Morrow, J.M.**, Miranda, M.A., Davagnanam, I., Cowley, P.C., Mehta, H., Hanna, M.G., Koltzenburg, M., Yousry, T.A., Reilly, M.M., et al. (2012a). Skeletal muscle MRI magnetisation transfer ratio reflects clinical severity in peripheral neuropathies. *J. Neurol. Neurosurg. Psychiatry* 83, 29–32.

Sinclair, C.D.J., **Morrow, J.M.**, Hanna, M.G., Reilly, M.M., Yousry, T.A., Golay, X., and Thornton, J.S. (2012b). Correcting radiofrequency inhomogeneity effects in skeletal muscle magnetisation transfer maps. *NMR Biomed.* 25, 262–270.

Willis, T.A., Hollingsworth, K.G., Coombs, A., Sveen, M.-L., Andersen, S., Stojkovic, T., Eagle, M., Mayhew, A., de Sousa, P.L., Dewar, L., **Morrow, J.M.**, et al. (2013). Quantitative Muscle MRI as an Assessment Tool for Monitoring Disease Progression in LGMD2I: A Multicentre Longitudinal Study. *PLoS One* 8, e70993.

Willis, T.A., Hollingsworth, K.G., Coombs, A., Sveen, M.-L., Andersen, S., Stojkovic, T., Eagle, M., Mayhew, A., de Sousa, P.L., Dewar, L., **Morrow, J.M.**, et al. (2014). Quantitative magnetic resonance imaging in limb-girdle muscular dystrophy 2I: a multinational cross-sectional study. *PLoS One* 9, e90377.

## 8 References

- Ababneh, Z.Q., Ababneh, R., Maier, S.E., Winalski, C.S., Oshio, K., Ababneh, A.M., and Mulkern, R.V. (2008). On the correlation between T(2) and tissue diffusion coefficients in exercised muscle: quantitative measurements at 3T within the tibialis anterior. *Magma N. Y. N* 21, 273–278.
- Agley, C.C., Rowlerson, A.M., Velloso, C.P., Lazarus, N.R., and Harridge, S.D.R. (2013). Human skeletal muscle fibroblasts, but not myogenic cells, readily undergo adipogenic differentiation. *J. Cell Sci.* 126, 5610–5625.
- Alanen, A.M., Falck, B., Kalimo, H., Komu, M.E., and Sonninen, V.H. (1994). Ultrasound, computed tomography and magnetic resonance imaging in myopathies: correlations with electromyography and histopathology. *Acta Neurol. Scand.* 89, 336–346.
- Amato, A.A., Sivakumar, K., Goyal, N., David, W.S., Salajegheh, M., Praestgaard, J., Lach-Trifilieff, E., Trendelenburg, A.-U., Laurent, D., Glass, D.J., et al. (2014). Treatment of sporadic inclusion body myositis with bimagrumab. *Neurology* 83, 2239–2246.
- Andersen, H., Gjerstad, M.D., and Jakobsen, J. (2004). Atrophy of foot muscles: a measure of diabetic neuropathy. *Diabetes Care* 27, 2382–2385.
- Andreassen, C.S., Jakobsen, J., Ringgaard, S., Ejksjaer, N., and Andersen, H. (2009). Accelerated atrophy of lower leg and foot muscles--a follow-up study of long-term diabetic polyneuropathy using magnetic resonance imaging (MRI). *Diabetologia* 52, 1182–1191.
- Arpan, I., Willcocks, R.J., Forbes, S.C., Finkel, R.S., Lott, D.J., Rooney, W.D., Triplett, W.T., Senesac, C.R., Daniels, M.J., Byrne, B.J., et al. (2014). Examination of effects of corticosteroids on skeletal muscles of boys with DMD using MRI and MRS. *Neurology* 83, 974–980.
- Barker, G.J., Tofts, P.S., and Gass, A. (1996). An interleaved sequence for accurate and reproducible clinical measurement of magnetization transfer ratio. *Magn. Reson. Imaging* 14, 403–411.
- Ben Saad, H., Préfaut, C., Tabka, Z., Zbidi, A., and Hayot, M. (2008). The forgotten message from gold: FVC is a primary clinical outcome measure of bronchodilator reversibility in COPD. *Pulm. Pharmacol. Ther.* 21, 767–773.

- Benveniste, O., Guiguet, M., Freebody, J., Dubourg, O., Squier, W., Maisonobe, T., Stojkovic, T., Leite, M.I., Allenbach, Y., Herson, S., et al. (2011). Long-term observational study of sporadic inclusion body myositis. *Brain J. Neurol.* 134, 3176–3184.
- Berciano, J., Gallardo, E., García, A., Ramón, C., Mateo, I., Infante, J., Rodríguez-Rodríguez, E., and Combarros, O. (2008). CMT1A duplication: refining the minimal adult phenotype. *J. Peripher. Nerv. Syst. JPNS* 13, 310–312.
- Berciano, J., Gallardo, E., García, A., Ramón, C., Infante, J., and Combarros, O. (2010). Clinical progression in Charcot-Marie-Tooth disease type 1A duplication: clinico-electrophysiological and MRI longitudinal study of a family. *J. Neurol.* 257, 1633–1641.
- Bland, J.M., and Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1, 307–310.
- BRATTON, C.B., HOPKINS, A.L., and WEINBERG, J.W. (1965). NUCLEAR MAGNETIC RESONANCE STUDIES OF LIVING MUSCLE. *Science* 147, 738–739.
- Bryan, W.W., Reisch, J.S., McDonald, G., Herbelin, L.L., Barohn, R.J., and Fleckenstein, J.L. (1998). Magnetic resonance imaging of muscle in amyotrophic lateral sclerosis. *Neurology* 51, 110–113.
- Budzik, J.F., Le Thuc, V., Demondion, X., Morel, M., Chechin, D., and Cotten, A. (2007). In vivo MR tractography of thigh muscles using diffusion imaging: initial results. *Eur. Radiol.* 17, 3079–3085.
- Burns, J., Ouvrier, R.A., Yiu, E.M., Joseph, P.D., Kornberg, A.J., Fahey, M.C., and Ryan, M.M. (2009). Ascorbic acid for Charcot-Marie-Tooth disease type 1A in children: a randomised, double-blind, placebo-controlled, safety and efficacy trial. *Lancet Neurol.* 8, 537–544.
- Bus, S.A., Yang, Q.X., Wang, J.H., Smith, M.B., Wunderlich, R., and Cavanagh, P.R. (2002). Intrinsic muscle atrophy and toe deformity in the diabetic neuropathic foot: a magnetic resonance imaging study. *Diabetes Care* 25, 1444–1450.
- Bushby, K., Finkel, R., Wong, B., Barohn, R., Campbell, C., Comi, G.P., Connolly, A.M., Day, J.W., Flanigan, K.M., Goemans, N., et al. (2014). Ataluren treatment of patients with nonsense mutation dystrophinopathy. *Muscle Nerve* 50, 477–487.

Bydder, G.M., Steiner, R.E., Young, I.R., Hall, A.S., Thomas, D.J., Marshall, J., Pallis, C.A., and Legg, N.J. (1982). Clinical NMR imaging of the brain: 140 cases. *AJR Am. J. Roentgenol.* 139, 215–236.

Cantwell, C., Ryan, M., O'Connell, M., Cunningham, P., Brennan, D., Costigan, D., Lynch, T., and Eustace, S. (2005). A comparison of inflammatory myopathies at whole-body turbo STIR MRI. *Clin. Radiol.* 60, 261–267.

Carlier, P.G. (2014). Global T2 versus water T2 in NMR imaging of fatty infiltrated muscles: different methodology, different information and different implications. *Neuromuscul. Disord. NMD* 24, 390–392.

Cermak, N.M., Noseworthy, M.D., Bourgeois, J.M., Tarnopolsky, M.A., and Gibala, M.J. (2012). Diffusion tensor MRI to assess skeletal muscle disruption following eccentric exercise. *Muscle Nerve* 46, 42–50.

de Certaines, J.D., Henriksen, O., Spisni, A., Cortsen, M., and Ring, P.B. (1993). In vivo measurements of proton relaxation times in human brain, liver, and skeletal muscle: a multicenter MRI study. *Magn. Reson. Imaging* 11, 841–850.

Chung, K.W., Suh, B.C., Shy, M.E., Cho, S.Y., Yoo, J.H., Park, S.W., Moon, H., Park, K.D., Choi, K.G., Kim, S., et al. (2008). Different clinical and magnetic resonance imaging features between Charcot-Marie-Tooth disease type 1A and 2A. *Neuromuscul. Disord. NMD* 18, 610–618.

Clark, B.C., Cook, S.B., and Ploutz-Snyder, L.L. (2007). Reliability of techniques to assess human neuromuscular function in vivo. *J. Electromyogr. Kinesiol. Off. J. Int. Soc. Electrophysiol. Kinesiol.* 17, 90–101.

Code of Federal Regulations Title 21, Volume 5, revised as of 2010 [Internet]. Available from:

<http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=314&showFR=1&subpartNode=21:5.0.1.1.4.8>

Commean, P.K., Tuttle, L.J., Hastings, M.K., Strube, M.J., and Mueller, M.J. (2011). Magnetic resonance imaging measurement reproducibility for calf muscle and adipose tissue volume. *J. Magn. Reson. Imaging JMRI* 34, 1285–1294.

Constantinides, C.D., Gillen, J.S., Boada, F.E., Pomper, M.G., and Bottomley, P.A. (2000). Human skeletal muscle: sodium MR imaging and quantification-potential applications in exercise and disease. *Radiology* 216, 559–568.

Cortese, A., Machado, P., Morrow, J., Dewar, L., Hiscock, A., Miller, A., Brady, S., Hilton-Jones, D., Parton, M., and Hanna, M.G. (2013). Longitudinal observational study of sporadic inclusion body myositis: implications for clinical trials. *Neuromuscul. Disord.* NMD 23, 404–412.

Coté, C., Hiba, B., Hebert, L.J., Vial, C., Remec, J.F., Janier, M., and Puymirat, J. (2011). MRI of tibialis anterior skeletal muscle in myotonic dystrophy type 1. *Can. J. Neurol. Sci. J. Can. Sci. Neurol.* 38, 112–118.

Cox, F.M., Reijnierse, M., van Rijswijk, C.S.P., Wintzen, A.R., Verschuuren, J.J., and Badrising, U.A. (2011a). Magnetic resonance imaging of skeletal muscles in sporadic inclusion body myositis. *Rheumatol. Oxf. Engl.* 50, 1153–1161.

Cox, F.M., Titulaer, M.J., Sont, J.K., Wintzen, A.R., Verschuuren, J.J.G.M., and Badrising, U.A. (2011b). A 12-year follow-up in sporadic inclusion body myositis: an end stage with major disabilities. *Brain J. Neurol.*

Cox, S., Limaye, V., Hill, C., Blumbergs, P., and Roberts-Thomson, P. (2010). Idiopathic inflammatory myopathies: diagnostic criteria, classification and epidemiological features. *Int. J. Rheum. Dis.* 13, 117–124.

Curiel, R.V., Jones, R., and Brindle, K. (2009). Magnetic resonance imaging of the idiopathic inflammatory myopathies: structural and clinical aspects. *Ann. N. Y. Acad. Sci.* 1154, 101–114.

Dalakas, M.C. (2010). Inflammatory muscle diseases: a critical review on pathogenesis and therapies. *Curr. Opin. Pharmacol.* 10, 346–352.

Dalakas, M.C., Sonies, B., Dambrosia, J., Sekul, E., Cupler, E., and Sivakumar, K. (1997). Treatment of inclusion-body myositis with IVIg: a double-blind, placebo-controlled study. *Neurology* 48, 712–716.

Damon, B.M., Froeling, M., Buck, A.K.W., Oudeman, J., Ding, Z., Nederveen, A.J., Bush, E.C., and Strijkers, G.J. (2016). Skeletal muscle diffusion tensor-MRI fiber tracking: rationale, data acquisition and analysis methods, applications and future directions. *NMR Biomed.*

De Gruttola, V., Fleming, T., Lin, D.Y., and Coombs, R. (1997). Perspective: validating surrogate markers--are we being naive? *J. Infect. Dis.* 175, 237–246.

Delmonico, M.J., Harris, T.B., Visser, M., Park, S.W., Conroy, M.B., Velasquez-Mieyer, P., Boudreau, R., Manini, T.M., Nevitt, M., Newman, A.B., et al. (2009). Longitudinal

study of muscle strength, quality, and adipose tissue infiltration123. *Am. J. Clin. Nutr.* 90, 1579–1585.

Deoni, S.C.L., Rutt, B.K., and Peters, T.M. (2003). Rapid combined T1 and T2 mapping using gradient recalled acquisition in the steady state. *Magn. Reson. Med. Off. J. Soc. Magn. Reson. Med. Soc. Magn. Reson. Med.* 49, 515–526.

Deroide, N., Bousson, V., Mambre, L., Vicaut, E., Laredo, J.D., and Kubis, N. (2015). Muscle MRI STIR signal intensity and atrophy are correlated to focal lower limb neuropathy severity. *Eur. Radiol.* 25, 644–651.

Deyo, R.A., Diehr, P., and Patrick, D.L. (1991). Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control. Clin. Trials* 12, 142S–158S.

Dimachkie, M., Machado, P., He, J., Wang, Y., McVey, A., Pasnoor, M., Gallagher, P., Herbelin, L., Statland, J., Miller, A., et al. (2014). Lack of Correlation Between Biomarker and Clinical Outcome Measures in the Arimoclomol Inclusion Body Myositis Pilot Study (S16.001). *Neurology* 82, S16.001.

Evans, M.R.B., Morrow, J.M., Sinclair, C.D.J., Shah, S., Reilly, M.M., Thornton, J.S., Yousry, T.A., and Hanna, M.G. (2015). Accurate Slice Selection Improves Responsiveness of Quantitative Lower Limb Muscle Mri in Cmt1a Patients. *Muscle Nerve* 52, S4–S4.

Finlayson, S., Morrow, J.M., Rodriguez Cruz, P.M., Sinclair, C.D.J., Fischmann, A., Thornton, J.S., Knight, S., Norbury, R., White, M., Al-Hajjar, M., et al. (2016). Muscle MRI in congenital myasthenic syndromes. *Muscle Nerve*.

Fischer, D., Kley, R.A., Strach, K., Meyer, C., Sommer, T., Eger, K., Rolfs, A., Meyer, W., Pou, A., Pradas, J., et al. (2008). Distinct muscle imaging patterns in myofibrillar myopathies. *Neurology* 71, 758–765.

Fischmann, A., Hafner, P., Fasler, S., Gloor, M., Bieri, O., Studler, U., and Fischer, D. (2012). Quantitative MRI can detect subclinical disease progression in muscular dystrophy. *J. Neurol.* 259, 1648–1654.

Fischmann, A., Morrow, J.M., Sinclair, C.D.J., Reilly, M.M., Hanna, M.G., Yousry, T., and Thornton, J.S. (2014). Improved anatomical reproducibility in quantitative lower-limb muscle MRI. *J. Magn. Reson. Imaging JMRI* 39, 1033–1038.

Fleming, T.R., and DeMets, D.L. (1996). Surrogate end points in clinical trials: are we being misled? *Ann. Intern. Med.* 125, 605–613.

Fontaine, B., Khurana, T.S., Hoffman, E.P., Bruns, G.A., Haines, J.L., Trofatter, J.A., Hanson, M.P., Rich, J., McFarlane, H., and Yasek, D.M. (1990). Hyperkalemic periodic paralysis and the adult muscle sodium channel alpha-subunit gene. *Science* 250, 1000–1002.

Forbes, S.C., Willcocks, R.J., Triplett, W.T., Rooney, W.D., Lott, D.J., Wang, D.-J., Pollaro, J., Senesac, C.R., Daniels, M.J., Finkel, R.S., et al. (2014). Magnetic resonance imaging and spectroscopy assessment of lower extremity skeletal muscles in boys with Duchenne muscular dystrophy: a multicenter cross sectional study. *PLoS One* 9, e106435.

Fortin, M., and Battié, M.C. (2012). Quantitative paraspinal muscle measurements: inter-software reliability and agreement using OsiriX and ImageJ. *Phys. Ther.* 92, 853–864.

Frisullo, G., Frusciante, R., Nociti, V., Tasca, G., Renna, R., Iorio, R., Patanella, A.K., Iannaccone, E., Marti, A., Rossi, M., et al. (2011). CD8(+) T cells in facioscapulohumeral muscular dystrophy patients with inflammatory features at muscle MRI. *J. Clin. Immunol.* 31, 155–166.

Gaeta, M., Scribano, E., Mileto, A., Mazziotti, S., Rodolico, C., Toscano, A., Settineri, N., Ascenti, G., and Blandino, A. (2011). Muscle fat fraction in neuromuscular disorders: dual-echo dual-flip-angle spoiled gradient-recalled MR imaging technique for quantification--a feasibility study. *Radiology* 259, 487–494.

Gaeta, M., Mileto, A., Mazzeo, A., Minutoli, F., Di Leo, R., Settineri, N., Donato, R., Ascenti, G., and Blandino, A. (2012). MRI findings, patterns of disease distribution, and muscle fat fraction calculation in five patients with Charcot-Marie-Tooth type 2 F disease. *Skeletal Radiol.* 41, 515–524.

Gallardo, E., García, A., Combarros, O., and Berciano, J. (2006). Charcot-Marie-Tooth disease type 1A duplication: spectrum of clinical and magnetic resonance imaging features in leg and foot muscles. *Brain J. Neurol.* 129, 426–437.

Garofalo, K., Penno, A., Schmidt, B.P., Lee, H.-J., Frosch, M.P., von Eckardstein, A., Brown, R.H., Hornemann, T., and Eichler, F.S. (2011). Oral L-serine supplementation reduces production of neurotoxic deoxysphingolipids in mice and humans with hereditary sensory autonomic neuropathy type 1. *J. Clin. Invest.* 121, 4735–4745.

- Gille, O., de Sèze, M.-P., Guérin, P., Jolivet, E., Vital, J.-M., and Skalli, W. (2011). Reliability of magnetic resonance imaging measurements of the cross-sectional area of the muscle contractile and non-contractile components. *Surg. Radiol. Anat. SRA* 33, 735–741.
- Gloor, M., Fasler, S., Fischmann, A., Haas, T., Bieri, O., Heinimann, K., Wetzel, S.G., Scheffler, K., and Fischer, D. (2011). Quantification of fat infiltration in oculopharyngeal muscular dystrophy: comparison of three MR imaging methods. *J. Magn. Reson. Imaging JMRI* 33, 203–210.
- Gloss, D., Moxley, R.T., Ashwal, S., and Oskoui, M. (2016). Practice guideline update summary: Corticosteroid treatment of Duchenne muscular dystrophy: Report of the Guideline Development Subcommittee of the American Academy of Neurology. *Neurology* 86, 465–472.
- Glover, G.H., and Schneider, E. (1991). Three-point Dixon technique for true water/fat decomposition with B<sub>0</sub> inhomogeneity correction. *Magn. Reson. Med. Off. J. Soc. Magn. Reson. Med. Soc. Magn. Reson. Med.* 18, 371–383.
- Guyatt, G., Walter, S., and Norman, G. (1987). Measuring change over time: Assessing the usefulness of evaluative instruments. *J. Chronic Dis.* 40, 171–178.
- Ha, D.-H., Choi, S., Kang, E.-J., and Park, H.T. (2015). Diffusion tensor imaging and T2 mapping in early denervated skeletal muscle in rats. *J. Magn. Reson. Imaging JMRI* 42, 617–623.
- Hatakenaka, M., Ueda, M., Ishigami, K., Otsuka, M., and Masuda, K. (2001). Effects of aging on muscle T2 relaxation time: difference between fast- and slow-twitch muscles. *Invest. Radiol.* 36, 692–698.
- Hatakenaka, M., Soeda, H., Okafuji, T., Yabuuchi, H., Shiokawa, S., Nishimura, J., and Honda, H. (2006). Steroid Myopathy: Evaluation of Fiber Atrophy with T2 Relaxation Time—Rabbit and Human Study<sup>1</sup>. *Radiology* 238, 650–657.
- Henriksen, O., de Certaines, J.D., Spisni, A., Cortsen, M., Muller, R.N., and Ring, P.B. (1993). In vivo field dependence of proton relaxation times in human brain, liver and skeletal muscle: a multicenter study. *Magn. Reson. Imaging* 11, 851–856.
- Hiba, B., Richard, N., Hébert, L.J., Coté, C., Nejjari, M., Vial, C., Bouhour, F., Puymirat, J., and Janier, M. (2012). Quantitative assessment of skeletal muscle degeneration in

patients with myotonic dystrophy type 1 using MRI. *J. Magn. Reson. Imaging JMRI* 35, 678–685.

Hilton-Jones, D., Miller, A., Parton, M., Holton, J., Sewry, C., and Hanna, M.G. (2010). Inclusion body myositis: MRC Centre for Neuromuscular Diseases, IBM workshop, London, 13 June 2008. *Neuromuscul. Disord.* 20, 142–147.

Hoffman, E.P., Brown Jr., R.H., and Kunkel, L.M. (1987). Dystrophin: The protein product of the duchenne muscular dystrophy locus. *Cell* 51, 919–928.

Hogrel, J.-Y., Wary, C., Moraux, A., Azzabou, N., Decostre, V., Ollivier, G., Canal, A., Lilien, C., Ledoux, I., Annoussamy, M., et al. (2016). Longitudinal functional and NMR assessment of upper limbs in Duchenne muscular dystrophy. *Neurology*.

Hollingsworth, K.G. (2015). Reducing acquisition time in clinical MRI by data undersampling and compressed sensing reconstruction. *Phys. Med. Biol.* 60, R297–322.

Holmbäck, A.M., Askaner, K., Holtås, S., Downham, D., and Lexell, J. (2002). Assessment of contractile and noncontractile components in human skeletal muscle by magnetic resonance imaging. *Muscle Nerve* 25, 251–258.

Horvath, J.J., Austin, S.L., Case, L.E., Greene, K.B., Jones, H.N., Soher, B.J., Kishnani, P.S., and Bashir, M.R. (2015). Correlation between quantitative whole-body muscle magnetic resonance imaging and clinical muscle weakness in Pompe disease. *Muscle Nerve* 51, 722–730.

Howald, H., Boesch, C., Kreis, R., Matter, S., Billeter, R., Essen-Gustavsson, B., and Hoppeler, H. (2002). Content of intramyocellular lipids derived by electron microscopy, biochemical assays, and (1)H-MR spectroscopy. *J. Appl. Physiol. Bethesda Md* 1985 92, 2264–2272.

Hoyte, L., Brubaker, L., Fielding, J.R., Lockhart, M.E., Heilbrun, M.E., Salomon, C.G., Ye, W., and Brown, M.B. (2009). Measurements from image-based three dimensional pelvic floor reconstruction: a study of inter- and intraobserver reliability. *J. Magn. Reson. Imaging JMRI* 30, 344–350.

Hoyte, L., Ye, W., Brubaker, L., Fielding, J.R., Lockhart, M.E., Heilbrun, M.E., Brown, M.B., and Warfield, S.K. (2011). Segmentations of MRI images of the female pelvic floor: a study of inter- and intra-reader reliability. *J. Magn. Reson. Imaging JMRI* 33, 684–691.

- Hu, Z.-J., He, J., Zhao, F.-D., Fang, X.-Q., Zhou, L.-N., and Fan, S.-W. (2011). An assessment of the intra- and inter-reliability of the lumbar paraspinal muscle parameters using CT scan and magnetic resonance imaging. *Spine* 36, E868-874.
- Huang, Y., Majumdar, S., Genant, H.K., Chan, W.P., Sharma, K.R., Yu, P., Mynhier, M., and Miller, R.G. (1994). Quantitative MR relaxometry study of muscle composition and function in Duchenne muscular dystrophy. *J. Magn. Reson. Imaging JMRI* 4, 59–64.
- Humbert, I.A., Reeder, S.B., Porcaro, E.J., Kays, S.A., Brittain, J.H., and Robbins, J. (2008). Simultaneous Estimation of Tongue Volume and Fat Fraction Using IDEAL-FSE. *J. Magn. Reson. Imaging JMRI* 28, 504–508.
- Jackson, C. e., Barohn, R. j., Gronseth, G., Pandya, S., and Herbelin, L. (2008). Inclusion body myositis functional rating scale: A reliable and valid measure of disease severity. *Muscle Nerve* 37, 473–476.
- Janiczek, R.L., Gambarota, G., Sinclair, C.D.J., Yousry, T.A., Thornton, J.S., Golay, X., and Newbould, R.D. (2011). Simultaneous T<sub>2</sub> and lipid quantitation using IDEAL-CPMG. *Magn. Reson. Med. Off. J. Soc. Magn. Reson. Med. Soc. Magn. Reson. Med.* 66, 1293–1302.
- Janssen, B.H., Voet, N.B.M., Nabuurs, C.I., Kan, H.E., de Rooy, J.W.J., Geurts, A.C., Padberg, G.W., van Engelen, B.G.M., and Heerschap, A. (2014). Distinct disease phases in muscles of facioscapulohumeral dystrophy patients identified by MR detected fat infiltration. *PloS One* 9, e85416.
- Jia, B.-X., Yang, Q., Li, S.-Y., Wan, M., Wang, H., Huo, L.-Y., Zhao, E., Ding, Y.-C., Ji, X.-M., and Guo, X.-H. (2015). Muscle edema of the lower limb determined by MRI in Asian hypokalaemic periodic paralysis patients. *Neurol. Res.* 37, 246–252.
- Johnson, M.A., Polgar, J., Weightman, D., and Appleton, D. (1973). Data on the distribution of fibre types in thirty-six human muscles. An autopsy study. *J. Neurol. Sci.* 18, 111–129.
- Jonas, D., Conrad, B., Von Einsiedel, H.G., and Bischoff, C. (2000). Correlation between quantitative EMG and muscle MRI in patients with axonal neuropathy. *Muscle Nerve* 23, 1265–1269.

- Jubrias, S.A., Odderson, I.R., Esselman, P.C., and Conley, K.E. (1997). Decline in isokinetic force with age: muscle cross-sectional area and specific force. *Pflüg. Arch. Eur. J. Physiol.* *434*, 246–253.
- Jurkat-Rott, K., Weber, M.-A., Fauler, M., Guo, X.-H., Holzherr, B.D., Paczulla, A., Nordsborg, N., Joechle, W., and Lehmann-Horn, F. (2009). K<sup>+</sup>-dependent paradoxical membrane depolarization and Na<sup>+</sup> overload, major and reversible contributors to weakness by ion channel leaks. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 4036–4041.
- Kan, H.E., Scheenen, T.W.J., Wohlgemuth, M., Klomp, D.W.J., van Loosbroek-Wagenmans, I., Padberg, G.W., and Heerschap, A. (2009). Quantitative MR imaging of individual muscle involvement in facioscapulohumeral muscular dystrophy. *Neuromuscul. Disord. NMD* *19*, 357–362.
- Kilgour, A.H.M., Subedi, D., Gray, C.D., Deary, I.J., Lawrie, S.M., Wardlaw, J.M., and Starr, J.M. (2012). Design and validation of a novel method to measure cross-sectional area of neck muscles included during routine MR brain volume imaging. *PLoS One* *7*, e34444.
- Kim, H.K., Laor, T., Horn, P.S., Racadio, J.M., Wong, B., and Dardzinski, B.J. (2010a). T2 mapping in Duchenne muscular dystrophy: distribution of disease activity and correlation with clinical assessments. *Radiology* *255*, 899–908.
- Kim, H.K., Laor, T., Horn, P.S., and Wong, B. (2010b). Quantitative assessment of the T2 relaxation time of the gluteus muscles in children with Duchenne muscular dystrophy: a comparative study before and after steroid treatment. *Korean J. Radiol. Off. J. Korean Radiol. Soc.* *11*, 304–311.
- Kim, S.M., Lee, J., Yoon, B.R., Kim, Y.J., Choi, B.-O., and Chung, K.W. (2015). Severe phenotypes in a Charcot-Marie-Tooth 1A patient with PMP22 triplication. *J. Hum. Genet.* *60*, 103–106.
- Kinali, M., Arechavala-Gomez, V., Cirak, S., Glover, A., Guglieri, M., Feng, L., Hollingsworth, K.G., Hunt, D., Jungbluth, H., Roper, H.P., et al. (2011). Muscle histology vs MRI in Duchenne muscular dystrophy. *Neurology* *76*, 346–353.
- Kirshner, B., and Guyatt, G. (1985). A methodological framework for assessing health indices. *J. Chronic Dis.* *38*, 27–36.
- Kishnani, P.S., Corzo, D., Nicolino, M., Byrne, B., Mandel, H., Hwu, W.L., Leslie, N., Levine, J., Spencer, C., McDonald, M., et al. (2007). Recombinant human acid [alpha]-

glucosidase: major clinical benefits in infantile-onset Pompe disease. *Neurology* 68, 99–109.

Koltzenburg, M., and Yousry, T. (2007). Magnetic resonance imaging of skeletal muscle. *Curr. Opin. Neurol.* 20, 595–599.

Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., and Laval-Jeantet, M. (1986). MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology* 161, 401–407.

Lehr, R. (1992). Sixteen S-squared over D-squared: a relation for crude sample size estimates. *Stat. Med.* 11, 1099–1102.

Lewis, R.A., McDermott, M.P., Herrmann, D.N., Hoke, A., Clawson, L.L., Siskind, C., Feely, S.M.E., Miller, L.J., Barohn, R.J., Smith, P., et al. (2013). High-dosage ascorbic acid treatment in Charcot-Marie-Tooth disease type 1A: results of a randomized, double-masked, controlled trial. *JAMA Neurol.* 70, 981–987.

Liang, M.H., Fossel, A.H., and Larson, M.G. (1990). Comparisons of five health status instruments for orthopedic evaluation. *Med. Care* 28, 632–642.

Longwei, X. (2012). Clinical application of diffusion tensor magnetic resonance imaging in skeletal muscle. *Muscles Ligaments Tendons J.* 2, 19–24.

Lunn, M.P.T., and Willison, H.J. (2009). Diagnosis and treatment in inflammatory neuropathies. *J. Neurol. Neurosurg. Psychiatry* 80, 249–258.

Lupski, J.R., de Oca-Luna, R.M., Slaugenhaupt, S., Pentao, L., Guzzetta, V., Trask, B.J., Saucedo-Cardenas, O., Barker, D.F., Killian, J.M., Garcia, C.A., et al. (1991). DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell* 66, 219–232.

Lynch, N.A., Metter, E.J., Lindle, R.S., Fozard, J.L., Tobin, J.D., Roy, T.A., Fleg, J.L., and Hurley, B.F. (1999). Muscle quality. I. Age-associated differences between arm and leg muscle groups. *J. Appl. Physiol.* Bethesda Md 1985 86, 188–194.

Maillard, S.M., Jones, R., Owens, C., Pilkington, C., Woo, P., Wedderburn, L.R., and Murray, K.J. (2004a). Quantitative assessment of MRI T2 relaxation time of thigh muscles in juvenile dermatomyositis. *Rheumatology* 43, 603–608.

Maillard, S.M., Jones, R., Owens, C., Pilkington, C., Woo, P., Wedderburn, L.R., and Murray, K.J. (2004b). Quantitative assessment of MRI T2 relaxation time of thigh muscles in juvenile dermatomyositis. *Rheumatol. Oxf. Engl.* 43, 603–608.

Maillard, S.M., Jones, R., Owens, C.M., Pilkington, C., Woo, P.M., Wedderburn, L.R., and Murray, K.J. (2005). Quantitative assessments of the effects of a single exercise session on muscles in juvenile dermatomyositis. *Arthritis Rheum.* 53, 558–564.

Manzur, A.Y., Kuntzer, T., Pike, M., and Swan, A. (2008). Glucocorticoid corticosteroids for Duchenne muscular dystrophy. *Cochrane Database Syst. Rev.* Online CD003725.

Maughan, R.J., Watson, J.S., and Weir, J. (1983). Strength and cross-sectional area of human skeletal muscle. *J. Physiol.* 338, 37–49.

McDaniel, J.D., Ulmer, J.L., Prost, R.W., Franczak, M.B., Jaradeh, S., Hamilton, C.A., and Mark, L.P. (1999). Magnetization transfer imaging of skeletal muscle in autosomal recessive limb girdle muscular dystrophy. *J. Comput. Assist. Tomogr.* 23, 609–614.

McDonald, C.M., Henricson, E.K., Han, J.J., Abresch, R.T., Nicorici, A., Atkinson, L., Elfring, G.L., Reha, A., and Miller, L.L. (2010). The 6-minute walk test in Duchenne/Becker muscular dystrophy: longitudinal observations. *Muscle Nerve* 42, 966–974.

McRobbie, D.W., Moore, E.A., Graves, M.J., and Prince, M.R. (2007). *MRI from Picture to Proton* (Cambridge University Press).

Meldrum, D., Cahalane, E., Conroy, R., Fitzgerald, D., and Hardiman, O. (2007). Maximum voluntary isometric contraction: reference values and clinical application. *Amyotroph. Lateral Scler. Off. Publ. World Fed. Neurol. Res. Group Mot. Neuron Dis.* 8, 47–55.

Mercuri, E., Talim, B., Moghadaszadeh, B., Petit, N., Brockington, M., Counsell, S., Guicheney, P., Muntoni, F., and Merlini, L. (2002a). Clinical and imaging findings in six cases of congenital muscular dystrophy with rigid spine syndrome linked to chromosome 1p (RSMD1). *Neuromuscul. Disord.* NMD 12, 631–638.

Mercuri, E., Cini, C., Counsell, S., Allsop, J., Zolkipli, Z., Jungbluth, H., Sewry, C., Brown, S.C., Pepe, G., and Muntoni, F. (2002b). Muscle MRI findings in a three-generation family affected by Bethlem myopathy. *Eur. J. Paediatr. Neurol. EJPN Off. J. Eur. Paediatr. Neurol. Soc.* 6, 309–314.

- Mercuri, E., Pichiecchio, A., Allsop, J., Messina, S., Pane, M., and Muntoni, F. (2007). Muscle MRI in inherited neuromuscular disorders: past, present, and future. *J. Magn. Reson. Imaging JMRI* 25, 433–440.
- Mercuri, E., Clements, E., Offiah, A., Pichiecchio, A., Vasco, G., Bianco, F., Berardinelli, A., Manzur, A., Pane, M., Messina, S., et al. (2010). Muscle magnetic resonance imaging involvement in muscular dystrophies with rigidity of the spine. *Ann. Neurol.* 67, 201–208.
- Micallef, J., Attarian, S., Dubourg, O., Gonnaud, P.-M., Hogrel, J.-Y., Stojkovic, T., Bernard, R., Jouve, E., Pitel, S., Vacherot, F., et al. (2009). Effect of ascorbic acid in patients with Charcot-Marie-Tooth disease type 1A: a multicentre, randomised, double-blind, placebo-controlled trial. *Lancet Neurol.* 8, 1103–1110.
- Miokovic, T., Armbrrecht, G., Felsenberg, D., and Belavy, D.L. (2012). Heterogeneous atrophy occurs within individual lower limb muscles during 60 days of bed rest. *J. Appl. Physiol. Bethesda Md* 1985 113, 1545–1559.
- Morrow, J.M., Matthews, E., Raja Rayan, D.L., Fischmann, A., Sinclair, C.D.J., Reilly, M.M., Thornton, J.S., Hanna, M.G., and Yousry, T.A. (2013). Muscle MRI reveals distinct abnormalities in genetically proven non-dystrophic myotonias. *Neuromuscul. Disord. NMD* 23, 637–646.
- Murphy, S.M., Davidson, G.L., Laura, M., Salih, M.A.M., Muntoni, F., Lunn, M.P.T., and Blake, J. (2011a). Reliability of the CMT Neuropathy Score (second version) in Charcot-Marie-Tooth Disease (abstract). *J. Peripher. Nerv. Syst.* 16, S93-4.
- Murphy, S.M., Herrmann, D.N., McDermott, M.P., Scherer, S.S., Shy, M.E., Reilly, M.M., and Pareyson, D. (2011b). Reliability of the CMT neuropathy score (second version) in Charcot-Marie-Tooth disease. *J. Peripher. Nerv. Syst. JPNS* 16, 191–198.
- Nakai, R., Azuma, T., Sudo, M., Urayama, S.-I., Takizawa, O., and Tsutsumi, S. (2008). MRI analysis of structural changes in skeletal muscles and surrounding tissues following long-term walking exercise with training equipment. *J. Appl. Physiol. Bethesda Md* 1985 105, 958–963.
- O'Brien, M.D. (2000). *Aids to the Examination of the Peripheral Nervous System* (Elsevier Health Sciences).

Ohana, M., Durand, M.-C., Marty, C., Lazareth, J.-P., Maisonobe, T., Mompoin, D., and Carlier, R.-Y. (2014). Whole-body muscle MRI to detect myopathies in non-extrapyramidal bent spine syndrome. *Skeletal Radiol.* *43*, 1113–1122.

Ortiz-Nieto, F., Johansson, L., Ahlström, H., and Weis, J. (2010). Quantification of lipids in human lower limbs using yellow bone marrow as the internal reference: gender-related effects. *Magn. Reson. Imaging* *28*, 676–682.

Pareyson, D., Reilly, M.M., Schenone, A., Fabrizi, G.M., Cavallaro, T., Santoro, L., Vita, G., Quattrone, A., Padua, L., Gemignani, F., et al. (2011a). Ascorbic acid in Charcot-Marie-Tooth disease type 1A (CMT-TRIAAL and CMT-TRAUK): a double-blind randomised trial. *Lancet Neurol.* *10*, 320–328.

Pareyson, D., Reilly, M.M., Schenone, A., Fabrizi, G.M., Cavallaro, T., and Santoro, L. (2011b). How to detect disease progression and treatment effect in Charcot-Marie-Tooth disease? Responsiveness of clinical outcome measures (Potomac, Maryland, USA).

Patzkó, Á., and Shy, M.E. (2010). Update on Charcot-Marie-Tooth Disease. *Curr. Neurol. Neurosci. Rep.* *11*, 78–88.

Pelayo-Negro, A.L., Gallardo, E., García, A., Sánchez-Juan, P., Infante, J., and Berciano, J. (2014). Evolution of Charcot-Marie-Tooth disease type 1A duplication: a 2-year clinico-electrophysiological and lower-limb muscle MRI longitudinal study. *J. Neurol.* *261*, 675–685.

Peng, A., Koffman, B.M., Malley, J.D., and Dalakas, M.C. (2000). Disease progression in sporadic inclusion body myositis: observations in 78 patients. *Neurology* *55*, 296–298.

Petrie, A., and Sabin, C. (2005). *Medical Statistics at a Glance* (Oxford, UK: Blackwell Publishing).

Pfarrmann, C.W.A., Schmid, M.R., Zanetti, M., Jost, B., Gerber, C., and Hodler, J. (2004). Assessment of fat content in supraspinatus muscle with proton MR spectroscopy in asymptomatic volunteers and patients with supraspinatus tendon lesions. *Radiology* *232*, 709–715.

Phillips, B.A., Cala, L.A., Thickbroom, G.W., Melsom, A., Zilko, P.J., and Mastaglia, F.L. (2001). Patterns of muscle involvement in inclusion body myositis: clinical and magnetic resonance imaging study. *Muscle Nerve* *24*, 1526–1534.

Phoenix, J., Betal, D., Roberts, N., Helliwell, T.R., and Edwards, R.H. (1996). Objective quantification of muscle and fat in human dystrophic muscle by magnetic resonance image analysis. *Muscle Nerve* 19, 302–310.

Pichiecchio, A., Poloni, G.U., Ravaglia, S., Ponzio, M., Germani, G., Maranzana, D., Costa, A., Repetto, A., Tavazzi, E., Danesino, C., et al. (2009). Enzyme replacement therapy in adult-onset glycogenosis II: is quantitative muscle MRI helpful? *Muscle Nerve* 40, 122–125.

van der Ploeg, A.T., Clemens, P.R., Corzo, D., Escolar, D.M., Florence, J., Groeneveld, G.J., Herson, S., Kishnani, P.S., Laforet, P., Lake, S.L., et al. (2010). A randomized study of alglucosidase alfa in late-onset Pompe's disease. *N. Engl. J. Med.* 362, 1396–1406.

Ploutz-Snyder, L.L., Nyren, S., Cooper, T.G., Potchen, E.J., and Meyer, R.A. (1997). Different effects of exercise and edema on T2 relaxation in skeletal muscle. *Magn. Reson. Med. Off. J. Soc. Magn. Reson. Med. Soc. Magn. Reson. Med.* 37, 676–682.

Ploutz-Snyder, L.L., Clark, B.C., Logan, L., and Turk, M. (2006). Evaluation of spastic muscle in stroke survivors using magnetic resonance imaging and resistance to passive motion. *Arch. Phys. Med. Rehabil.* 87, 1636–1642.

Prentice, R.L. (1989). Surrogate endpoints in clinical trials: definition and operational criteria. *Stat. Med.* 8, 431–440.

Psatha, M., Wu, Z., Gammie, F.M., Ratkevicius, A., Wackerhage, H., Lee, J.H., Redpath, T.W., Gilbert, F.J., Ashcroft, G.P., Meakin, J.R., et al. (2012). A longitudinal MRI study of muscle atrophy during lower leg immobilization following ankle fracture. *J. Magn. Reson. Imaging JMRI* 35, 686–695.

Quijano-Roy, S., Avila-Smirnow, D., Carlier, R.Y., and WB-MRI muscle study group (2012). Whole body muscle MRI protocol: pattern recognition in early onset NM disorders. *Neuromuscul. Disord. NMD* 22 *Suppl* 2, S68-84.

Raeymaekers, P., Timmerman, V., Nelis, E., De Jonghe, P., Hoogendijk, J.E., Baas, F., Barker, D.F., Martin, J.J., De Visser, M., and Bolhuis, P.A. (1991). Duplication in chromosome 17p11.2 in Charcot-Marie-Tooth neuropathy type 1a (CMT 1a). The HMSN Collaborative Research Group. *Neuromuscul. Disord. NMD* 1, 93–97.

Raja Rayan, D.L., and Hanna, M.G. (2010). Skeletal muscle channelopathies: nondystrophic myotonias and periodic paralysis. *Curr. Opin. Neurol.* 23, 466–476.

- Ranson, C.A., Burnett, A.F., Kerslake, R., Batt, M.E., and O'Sullivan, P.B. (2006). An investigation into the use of MR imaging to determine the functional cross sectional area of lumbar paraspinal muscles. *Eur. Spine J. Off. Publ. Eur. Spine Soc. Eur. Spinal Deform. Soc. Eur. Sect. Cerv. Spine Res. Soc.* 15, 764–773.
- Ravaglia, S., Pichiecchio, A., Ponzio, M., Danesino, C., Saeidi Garaghani, K., Poloni, G.U., Toscano, A., Moglia, A., Carlucci, A., Bini, P., et al. (2010). Changes in skeletal muscle qualities during enzyme replacement therapy in late-onset type II glycogenosis: temporal and spatial pattern of mass vs. strength response. *J. Inherit. Metab. Dis.* 33, 737–745.
- Reeder, S.B., Hu, H.H., and Sirlin, C.B. (2012). Proton density fat-fraction: A standardized mr-based biomarker of tissue fat concentration. *J. Magn. Reson. Imaging* 36, 1011–1014.
- Reilly, M.M., de Jonghe, P., and Pareyson, D. (2006). 136th ENMC International Workshop: Charcot-Marie-Tooth disease type 1A (CMT1A)8-10 April 2005, Naarden, The Netherlands. *Neuromuscul. Disord. NMD* 16, 396–402.
- Reilly, M.M., Murphy, S.M., and Laurá, M. (2011). Charcot-Marie-Tooth disease. *J. Peripher. Nerv. Syst. JPNS* 16, 1–14.
- Reimers, C.D., Schedel, H., Fleckenstein, J.L., Nägele, M., Witt, T.N., Pongratz, D.E., and Vogl, T.J. (1994). Magnetic resonance imaging of skeletal muscles in idiopathic inflammatory myopathies of adults. *J. Neurol.* 241, 306–314.
- Ricotti, V., Evans, R.B., Sinclair, C.D.J., Morrow, J.M., Butler, J.W., Janiczek, R.L., Hanna, M.G., Matthews, P.M., Yousry, T.A., Muntoni, F., et al. (2014). Upper limb muscle MRI fat-water quantification and clinical functional correlation in non-ambulant Duchenne muscular dystrophy. *Neuromuscul. Disord.* 24, 839–839.
- Roach, K.E. (2006). Measurement of Health Outcomes: Reliability, Validity and Responsiveness. *J. Prosthet. Orthot.* 18, 8–12.
- Robinson, L.R., Temkin, N.R., Fujimoto, W.Y., and Stolov, W.C. (1991). Effect of statistical methodology on normal limits in nerve conduction studies. *Muscle Nerve* 14, 1084–1090.
- Sahraian, M.A., and Eshaghi, A. (2010). Role of MRI in diagnosis and treatment of multiple sclerosis. *Clin. Neurol. Neurosurg.* 112, 609–615.

- Sakurai, Y., Tamura, Y., Takeno, K., Sato, F., Fujitani, Y., Hirose, T., Kawamori, R., and Watada, H. (2011). Association of T2 relaxation time determined by magnetic resonance imaging and intramyocellular lipid content of the soleus muscle in healthy subjects. *J. Diabetes Investig.* 2, 356–358.
- Saupe, N., White, L.M., Sussman, M.S., Kassner, A., Tomlinson, G., and Noseworthy, M.D. (2008). Diffusion tensor magnetic resonance imaging of the human calf: comparison between 1.5 T and 3.0 T-preliminary results. *Invest. Radiol.* 43, 612–618.
- Schick, F., Machann, J., Brechtel, K., Strempler, A., Klumpp, B., Stein, D.T., and Jacob, S. (2002). MRI of muscular fat. *Magn. Reson. Med. Off. J. Soc. Magn. Reson. Med. Soc. Magn. Reson. Med.* 47, 720–727.
- Schwenzer, N.F., Martirosian, P., Machann, J., Schraml, C., Steidle, G., Claussen, C.D., and Schick, F. (2009). Aging effects on human calf muscle properties assessed by MRI at 3 Tesla. *J. Magn. Reson. Imaging JMRI* 29, 1346–1354.
- Sekul, E.A., Chow, C., and Dalakas, M.C. (1997). Magnetic resonance imaging of the forearm as a diagnostic aid in patients with sporadic inclusion body myositis. *Neurology* 48, 863–866.
- Shen, W., Mao, X., Wolper, C., Heshka, S., Dashnaw, S., Hirsch, J., Heymsfield, S.B., and Shungu, D.C. (2008). Reproducibility of single- and multi-voxel <sup>1</sup>H MRS measurements of intramyocellular lipid in overweight and lean subjects under conditions of controlled dietary calorie and fat intake. *NMR Biomed.* 21, 498–506.
- Shy, M.E., Blake, J., Krajewski, K., Fuerst, D.R., Laura, M., Hahn, A.F., Li, J., Lewis, R.A., and Reilly, M. (2005). Reliability and validity of the CMT neuropathy score as a measure of disability. *Neurology* 64, 1209–1214.
- Shy, M.E., Chen, L., Swan, E.R., Taube, R., Krajewski, K.M., Herrmann, D., Lewis, R.A., and McDermott, M.P. (2008). Neuropathy progression in Charcot-Marie-Tooth disease type 1A. *Neurology* 70, 378–383.
- Sinclair, C.D.J., Samson, R.S., Thomas, D.L., Weiskopf, N., Lutti, A., Thornton, J.S., and Golay, X. (2010). Quantitative magnetization transfer in in vivo healthy human skeletal muscle at 3 T. *Magn. Reson. Med. Off. J. Soc. Magn. Reson. Med. Soc. Magn. Reson. Med.* 64, 1739–1748.
- Sinclair, C.D.J., Morrow, J.M., Miranda, M.A., Davagnanam, I., Cowley, P.C., Mehta, H., Hanna, M.G., Koltzenburg, M., Yousry, T.A., Reilly, M.M., et al. (2012a). Skeletal

muscle MRI magnetisation transfer ratio reflects clinical severity in peripheral neuropathies. *J. Neurol. Neurosurg. Psychiatry* 83, 29–32.

Sinclair, C.D.J., Morrow, J.M., Hanna, M.G., Reilly, M.M., Yousry, T.A., Golay, X., and Thornton, J.S. (2012b). Correcting radiofrequency inhomogeneity effects in skeletal muscle magnetisation transfer maps. *NMR Biomed.* 25, 262–270.

Sinha, S., Sinha, U., and Edgerton, V.R. (2006). In vivo diffusion tensor imaging of the human calf muscle. *J. Magn. Reson. Imaging JMRI* 24, 182–190.

Sinha, U., Csapo, R., Malis, V., Xue, Y., and Sinha, S. (2015). Age-related differences in diffusion tensor indices and fiber architecture in the medial and lateral gastrocnemius. *J. Magn. Reson. Imaging JMRI* 41, 941–953.

Smeulders, M.J.C., van den Berg, S., Oudeman, J., Nederveen, A.J., Kreulen, M., and Maas, M. (2010). Reliability of in vivo determination of forearm muscle volume using 3.0 T magnetic resonance imaging. *J. Magn. Reson. Imaging JMRI* 31, 1252–1255.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 23 *Suppl 1*, S208-219.

Solari, A., Laurà, M., Salsano, E., Radice, D., and Pareyson, D. (2008). Reliability of clinical outcome measures in Charcot-Marie-Tooth disease. *Neuromuscul. Disord.* NMD 18, 19–26.

Spector, S.A., Lemmer, J.T., Koffman, B.M., Fleisher, T.A., Feuerstein, I.M., Hurley, B.F., and Dalakas, M.C. (1997). Safety and efficacy of strength training in patients with sporadic inclusion body myositis. *Muscle Nerve* 20, 1242–1248.

Springer, I., Müller, M., Hamm, B., and Dewey, M. (2011). Intra- and interobserver variability of magnetic resonance imaging for quantitative assessment of abductor and external rotator muscle changes after total hip arthroplasty. *Eur. J. Radiol.*

Sproule, D.M., Punyanitya, M., Shen, W., Dashnaw, S., Martens, B., Montgomery, M., Montes, J., Battista, V., Finkel, R., Darras, B., et al. (2011a). Muscle volume estimation by magnetic resonance imaging in spinal muscular atrophy. *J. Child Neurol.* 26, 309–317.

Sproule, D.M., Montgomery, M.J., Punyanitya, M., Shen, W., Dashnaw, S., Montes, J., Dunaway, S., Finkel, R., Darras, B., De Vivo, D.C., et al. (2011b). Thigh Muscle

Volume Measured by Magnetic Resonance Imaging Is Stable Over a 6-Month Interval in Spinal Muscular Atrophy. *J. Child Neurol.*

Steidle, G., and Schick, F. (2015). Addressing spontaneous signal voids in repetitive single-shot DWI of musculature: spatial and temporal patterns in the calves of healthy volunteers and consideration of unintended muscle activities as underlying mechanism. *NMR Biomed.* 28, 801–810.

Stollberger, R., and Wach, P. (1996). Imaging of the active B1 field in vivo. *Magn. Reson. Med. Off. J. Soc. Magn. Reson. Med. Soc. Magn. Reson. Med.* 35, 246–251.

Streiner, D.L., and Norman, G.R. (2008). *Health measurement scales: a practical guide to their development and use* (Oxford, UK: Oxford University Press).

Tamura, Y., Watada, H., Igarashi, Y., Nomiya, T., Onishi, T., Takahashi, K., Doi, S., Katamoto, S., Hirose, T., Tanaka, Y., et al. (2008). Short-term effects of dietary fat on intramyocellular lipid in sprinters and endurance runners. *Metabolism.* 57, 373–379.

Tingart, M.J., Apreleva, M., Lehtinen, J.T., Capell, B., Palmer, W.E., and Warner, J.J.P. (2003). Magnetic resonance imaging in quantitative analysis of rotator cuff muscle volume. *Clin. Orthop.* 104–110.

Tollbäck, A., Söderlund, V., Jakobsson, F., Fransson, A., Borg, K., and Borg, J. (1996). Magnetic resonance imaging of lower extremity muscles and isokinetic strength in foot dorsiflexors in patients with prior polio. *Scand. J. Rehabil. Med.* 28, 115–123.

Torriani, M., Townsend, E., Thomas, B.J., Bredella, M.A., Ghomi, R.H., and Tseng, B.S. (2012). Lower leg muscle involvement in Duchenne muscular dystrophy: an MR imaging and spectroscopy study. *Skeletal Radiol.* 41, 437–445.

Triplett, W.T., Baligand, C., Forbes, S.C., Willcocks, R.J., Lott, D.J., DeVos, S., Pollaro, J., Rooney, W.D., Sweeney, H.L., Bönnemann, C.G., et al. (2014). Chemical shift-based MRI to measure fat fractions in dystrophic skeletal muscle. *Magn. Reson. Med. Off. J. Soc. Magn. Reson. Med. Soc. Magn. Reson. Med.* 72, 8–19.

Ulmer, J.L., Logani, S.C., Mark, L.P., Hamilton, C.A., Prost, R.W., and Garman, J.N. (1998). Near-resonance saturation pulse imaging of the extraocular muscles in thyroid-related ophthalmopathy. *AJNR Am. J. Neuroradiol.* 19, 943–950.

Vanhoutte, E.K., Faber, C.G., van Nes, S.I., Jacobs, B.C., van Doorn, P.A., and van Koningsveld, R. (2011). Is it time to change manual muscle testing: modifying the

Medical Research Council Grading System through Rasch analysis (abstract). *J. Peripher. Nerv. Syst.* 16, S139–S140.

Verhamme, C., van Schaik, I.N., Koelman, J.H.T.M., de Haan, R.J., and de Visser, M. (2009). The natural history of Charcot-Marie-Tooth type 1A in adults: a 5-year follow-up study. *Brain J. Neurol.* 132, 3252–3262.

Voronov, A.V. (2003). [Anatomical cross-sectional areas and volumes of the lower extremity muscles]. *Fiziol. Cheloveka* 29, 81–91.

Walter, M.C., Lochmüller, H., Toepfer, M., Schlotter, B., Reilich, P., Schröder, M., Müller-Felber, W., and Pongratz, D. (2000). High-dose immunoglobulin therapy in sporadic inclusion body myositis: a double-blind, placebo-controlled study. *J. Neurol.* 247, 22–28.

Ware, J.E., Jr, and Sherbourne, C.D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med. Care* 30, 473–483.

Wattjes, M.P., Kley, R.A., and Fischer, D. (2010). Neuromuscular imaging in inherited muscle diseases. *Eur Radiol* 20, 2447–2460.

Wessig, C., Koltzenburg, M., Reiners, K., Solymosi, L., and Bendszus, M. (2004). Muscle magnetic resonance imaging of denervation and reinnervation: correlation with electrophysiology and histology. *Exp. Neurol.* 185, 254–261.

Willcocks, R.J., Arpan, I.A., Forbes, S.C., Lott, D.J., Senesac, C.R., Senesac, E., Deol, J., Triplett, W.T., Baligand, C., Daniels, M.J., et al. (2014). Longitudinal measurements of MRI-T2 in boys with Duchenne muscular dystrophy: effects of age and disease progression. *Neuromuscul. Disord. NMD* 24, 393–401.

Willis, T.A., Hollingsworth, K.G., Coombs, A., Sveen, M.-L., Andersen, S., Stojkovic, T., Eagle, M., Mayhew, A., de Sousa, P.L., Dewar, L., et al. (2013). Quantitative Muscle MRI as an Assessment Tool for Monitoring Disease Progression in LGMD2I: A Multicentre Longitudinal Study. *PLoS One* 8, e70993.

Willis, T.A., Hollingsworth, K.G., Coombs, A., Sveen, M.-L., Andersen, S., Stojkovic, T., Eagle, M., Mayhew, A., de Sousa, P.L., Dewar, L., et al. (2014). Quantitative magnetic resonance imaging in limb-girdle muscular dystrophy 2I: a multinational cross-sectional study. *PLoS One* 9, e90377.

Wittes, J., Lakatos, E., and Probstfield, J. (1989). Surrogate endpoints in clinical trials: cardiovascular diseases. *Stat. Med.* 8, 415–425.

Wokke, B.H., Bos, C., Reijnierse, M., van Rijswijk, C.S., Eggers, H., Webb, A., Verschuuren, J.J., and Kan, H.E. (2013). Comparison of dixon and T1-weighted MR methods to assess the degree of fat infiltration in duchenne muscular dystrophy patients. *J. Magn. Reson. Imaging JMRI* 38, 619–624.

Wokke, B.H., van den Bergen, J.C., Versluis, M.J., Niks, E.H., Milles, J., Webb, A.G., van Zwet, E.W., Aartsma-Rus, A., Verschuuren, J.J., and Kan, H.E. (2014). Quantitative MRI and strength measurements in the assessment of muscle quality in Duchenne muscular dystrophy. *Neuromuscul. Disord. NMD* 24, 409–416.

Wood, L.E., Dixon, S., Grant, C., and Armstrong, N. (2004). Elbow flexion and extension strength relative to body or muscle size in children. *Med. Sci. Sports Exerc.* 36, 1977–1984.

Wren, T.A.L., Bluml, S., Tseng-Ong, L., and Gilsanz, V. (2008). Three-point technique of fat quantification of muscle tissue as a marker of disease progression in Duchenne muscular dystrophy: preliminary study. *AJR Am. J. Roentgenol.* 190, W8-12.

Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., and Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *NeuroImage* 31, 1116–1128.