Prognostic Indicators of Successful Rehabilitation Outcome in Patients with Subacromial Impingement Syndrome / Rotator Cuff Tendinopathy

Author:

Michael Smith BSc (Hons), MSc (Physiotherapy), PgC (Medical Ultrasound)

Supervisors:

Dr Valerie Sparkes, Prof. Robert van Deursen, Dr Nick Courtier

A thesis submitted for the degree of Doctor of Philosophy

Year of Presentation: 2018

School of Healthcare Sciences, Cardiff University

Cardiff, Wales, UK

Abstract

Background

As one of the most common types of shoulder pain, subacromial impingement / rotator cuff tendinopathy (SIS/RCTendinopathy) is frequently treated with physiotherapy. However there is uncertainty as to which patients do and do not improve with this approach.

Aim

To identify baseline factors that predict outcome in SIS/RCTendinopathy patients following treatment with physiotherapy.

Method

Patients with a clinical diagnosis of SIS/RCTendinopathy referred for physiotherapy were recruited. Baseline potential prognostic factors (demographics, clinical history, patient reported measures, clinical measures and structural pathology) were recorded immediately prior to the first physiotherapy appointment. Treatment was pragmatic as determined by the treating clinician. Outcome was determined by the change in a patient reported outcome measure (Oxford Shoulder Score, OSS) from baseline. Time points for analysis were at discharge and three months post-discharge.

Findings

Seventy-six patients fulfilling the inclusion and exclusion criteria were recruited and outcome data were available from 73 subjects at discharge and 62 subjects at three months post discharge. The number of candidate variables were trimmed using a sequential decision process comprising methodological, conceptual and statistical methods. Multivariate regression analysis comprising forward selection and backward elimination was applied. Baseline function (measured by the total SPADI; higher SPADI = greater pain and disability) and (younger) age predicted 15.7% of the outcome (greater improvement in function) from baseline to discharge. Total SPADI (higher SPADI = greater pain and disability) predicted 9.6% of the outcome (greater improvement in function) from baseline to three months post-discharge.

Conclusion

The study findings provide evidence of a limited ability to predict outcome. Potential reasons for this include the multifactorial nature of the condition and the high degree of heterogeneity in the treatment of the cohort. The limited sample size and lack of consideration of variable types such as

patient or clinician expectation are also noteworthy limitations of the study. Nonetheless, the regression findings indicate the value of considering baseline function and age for informing the likely change in patient reported pain and disability across the period of treatment and up to three months post discharge.

Acknowledgements

I am very grateful to Arthritis Research UK for their support of this project through their Nurse and Allied Health Professional Training Fellowship scheme. I would particularly like to thank Dr Claire Farmer and Dr Sarah George for their support and counsel.

I would like to thank my Supervisors Dr Valerie Sparkes, Professor Robert van Deursen and Dr Nicholas Courtier – and my external advisor Professor Karen Ginn – for their input at key points within this project. I would also like to thank the Arthritis Research UK Biomechanics and Bioengineering Centre at Cardiff for their support, particularly Dr Helen Hodgson and Dr Andrea Longman. I am also grateful for the support of the Department of Physiotherapy / School of Healthcare Sciences.

I would like to thank Steve Hiles who supported the project in his then capacity as a NISCHR Research Officer. I am very grateful for Steve's practical input, instruction on a variety of methodological and quality control mechanisms as well as his company and friendship at some challenging times.

I am grateful to the clinical staff in Cardiff and Vale Radiology and Physiotherapy departments (UHW and Whitchurch) for their practical support with the project. I am particularly grateful to the site leads, Anne Wilkes and Rob Goldsmith.

I would like to thank Alison Hall for her guidance, support and encouragement in my ultrasound career.

I would like to thank the numerous colleagues who have been the bearers of sanity and welcome distraction – Richard Day, Amy Lake, Katy Hamana, Una Jones, Nicola Gale, Arshi Iqbal, Kate Jones, Becci Hemming and Rob Letchford.

I would like to thank Dr Quaite and Dr Lawton for their professional support and to my beautiful family for being the light at the end of the tunnel.

I would like to dedicate this thesis to Alison Rogers who inspired and supported me at critical points in my ultrasound training, prior to her untimely death from breast cancer. I would also like to dedicate this thesis to my Dad who died just before submission. Rest in peace x

Contents

Table of tables

Table of figures

List of abbreviations

4DSQ	Four-dimensional symptom questionnaire
ACJ	Acromioclavicular Joint
AROM	Active range of movement
BMI	Body Mass Index
C&V UHB	Cardiff and Vale University Hospital Board
CAT	Computerized Adaptive Test
DASH	Disabilities of the Arm, Shoulder and Hand Score
DGH	District General Hospital
DNA	Did not attend
EQ-VAS	EuroQol-visual analogue scale
ER	External Rotation
FABQ	Fear Avoidance Beliefs Questionnaire
FLEXSF	Flexilevel Scale of Shoulder Function
GHJ	Glenohumeral Joint
GIRD	Glenohumeral Internal Rotation Deficit
GP	General Practitioner
GROC	Global Rating of Change Scale
HEP	Home exercise programme
HHD	Hand-held dynamometer
ICC	Intra-class Correlation
ICD	International Classification of Diseases
ICF	International Classification of Functioning
IR	Internal Rotation
Kg	Kilogram
KN	Kevin Nicholas
MCIC	Minimally Clinically Important Change
MCS	Mental Component Score
MS	Michael Smith
OSS	Oxford Shoulder Score
PCS	Physical Component Score
PgC	Postgraduate Certificate
PM	Dr Peter Mullaney
PST	Posterior Shoulder Tightness
RCT	Randomised Control Trial
RCTendinopathy	Rotator Cuff Tendinopathy
ROM	Range of Movement
SDT	Scapular Dyskinesis Test
SF-36	Short-Form Health Survey
SH	Steve Hiles
SIS	Subacromial Impingement Syndrome
SPADI	Shoulder Pain and Disability Index
SPADI-F	Functional subscale of SPADI
SPSS	Statistical Package for the Social Sciences

SST	Simple Shoulder Test
TSK	Tampa Scale for Kinesiophobia
UCLA	University of California Los Angeles
UHW	University Hospital Wales
UTA	Unable to attend
VAS	Visual Analogue Scale
VIF	Variance inflation factor
WHI	Whitchurch hospital

1 INTRODUCTION

1.1 The relevance of shoulder pain and its associated burden

Shoulder pain affects around 1 in 3 individuals during their lifetime (van der Heijden 1999), with reports of an annual prevalence of 41.2 to 48.4 per 1000 person-years (Greving et al. 2012) and point prevalence reported between 7% and 26% (Luime et al. 2004). Furthermore high rates of recurrence are reported (Luime et al. 2004). Together these reflect a high burden in the form of General Practitioner (GP) consultations and contact with treatment providers such as physiotherapy.

Shoulder pain can impact on upper limb function due to an inability or reticence to use the limb for activities of daily living, recreational or occupational tasks (Dong et al. 2015). With activities of daily living, the impact can therefore be on the individual (Gartsman et al. 1998) and those who the person cares for or is cared for by. With recreational activities, this can impact on the ability of the patient to participate; and therefore negatively influence the social, physical and mental health benefits often derived from such activities. From an occupational perspective, it can impact on the patient's ability to remain in employment which in turn can have economic, social and mental health consequences (Linaker and Walker-Bone 2015). Similarly, the impact can be upon employers and wider society (Holtermann et al. 2010). Optimising the management of patients with shoulder pain is therefore a high priority for the individual, healthcare providers and society.

1.2 Shoulder pain, Subacromial Impingement Syndrome and Rotator Cuff Tendinopathy

Shoulder pain can be of a variety of sources, including referral from the viscera and spinal structures into the shoulder (Linaker and Walker-Bone 2015). Separately, numerous categories of local shoulder pain are described, of which Subacromial Impingement Syndrome (SIS) and Rotator Cuff Tendinopathy (RCTendinopathy) are reported as being the most commonly assigned clinical diagnoses (van der Windt et al. 1995; Ostor et al. 2005). However there is widespread controversy in the literature as to the aetiology of SIS and RCTendinopathy, including what structures are at fault and how best to identify the pathology.

In light of this, alternative approaches to describing such patients have been proposed that focus more on the clinical presentation of shoulder disorders rather than the specific theorised aetiological

processes (Lewis 2009). Given the lack of consensus as to a definitive approach to categorising such patients and the likely overlap between SIS and RCTendinopathy (van der Windt et al. 1995; Jia et al. 2009; Hanchard et al. 2013), use of SIS/RCTendinopathy as a description of such patients has merit.

1.3 Treatment approaches including physiotherapy for patients with SIS/RCTendinopathy

The treatment approaches used for patients with SIS/RCTendinopathy can take many forms (Dong et al. 2015) but uncertainty persists as to which patients respond best to what treatment approaches and what aspects of the available care pathway. Indeed the uncertain nature of SIS/RCTendinopathy combined with the myriad treatment approaches available indicate that uncertainty remains regarding the relevance of underlying pathological processes to patient symptoms and their optimal management.

Physiotherapy is one of several treatment approaches that may be used and may include approaches such as education and advice, exercise based therapy and hands-on techniques such as manual therapy, taping and electrotherapy. The content of physiotherapy can vary greatly and may reflect the experience and preference of the clinician, the presenting nature of the shoulder complaint or patient preference (Kromer et al. 2013).

Evidence of the effectiveness of physiotherapy for SIS/RCTendinopathy is limited (Saltychev et al. 2015). Epidemiological, diagnostic and trial data (Green et al. 2003) can be viewed as providing evidence in this area although uncertainty persists as to which patients do and do not respond to different treatment approaches. One methodological approach that could shed light on this is prognostic research (Dinant et al. 2007).

1.4 The role prognostic research can play in informing clinical management of musculoskeletal disorders

Prognostic research is concerned with the identification of factors that predict a particular outcome (Moons et al. 2009a). Recognising the merit of developing targeted interventions (Selfe et al. 2016), a prognostic approach has the advantage of not being constrained by the requirements of diagnosis *per se*. Instead this approach allows for the potential inclusion of multiple and often disparate factors to explore their individual and collective influence on outcome (Royston et al. 2009).

Whilst prognostic research has been published across different medical specialities, including rehabilitation, musculoskeletal disorders and shoulder pain, the robustness and relative value of these studies has been variable (Hemingway et al. 2013). This has led researchers to highlight the limited value of some previously published prognostic research and thus the importance of establishing higher standards with the undertaking, reporting and interpreting of prognostic research (Kent et al. 2010).

Key factors to consider here include establishing the reliability of the measurement tools used, incorporating clinically applicable measurement approaches, ensuring alignment between the number of variables and the sample size and using a clinically relevant timescale for follow up (Altman et al. 2009). Alongside this, the viewing of prognostic research within a method framework in order to reflect the nature of the evidence provided has been advocated. This includes exploratory hypothesis-setting studies, hypothesis-testing (confirmatory) studies in independent samples and subsequent validation studies (Kent et al. 2010).

This thesis will therefore seek to address some of the above methodological concerns regarding prognostic research in exploring predictors of outcome in patients with SIS/RCTendinopathy treated with physiotherapy.

2 LITERATURE REVIEW

2.1 Overview

The following is a critical review of the relevant literature relating to this study. Conceptual frameworks relating to subacromial impingement and rotator cuff tendinopathy will firstly be presented. The strategy by which the prognostic studies to be critiqued were identified will then be described. Key aspects of study design, pathology, patient care pathway and sample will be explored, leading onto a consideration of the potential prognostic factors considered by previous studies. The treatment approaches will be discussed along with how outcome can be defined. The prognostic analysis approaches used and the prognostic findings for each variable type and study will be presented along with a consideration of the strength of evidence. These will culminate in the study to be undertaken.

Recent systematic reviews in this area have looked at response to physiotherapy treatment for musculoskeletal shoulder pain (Chester et al. 2013) and adults undergoing physical therapy for rotator cuff disorders (Braun et al. 2016). As with the previous systematic reviews in this area by van der Heijden (1999) and Kuijpers et al. (2004), high levels of study heterogeneity were identified and so all previous reviews have undertaken narrative analyses. The most recent review (Braun et al. 2016) emphasised the short-comings of attempting to synthesise evidence regarding individual prognostic factors rather than actual prognostic models. In addition they identified all studies in their review as being at high risk of bias when assessed by the 'Prediction Study Risk of Bias Assessment Tool; PROBAST' (Wolff (2017) personal communication).

2.2 Conceptual frameworks relating to subacromial impingement and rotator cuff tendinopathy

Various theoretical models of subacromial impingement and rotator cuff tendinopathy disease have been proposed in the literature. These provide parallel frameworks for the diagnosis, treatment and potentially prognosis of SIS/RCTendinopathy.

2.2.1 Patho-anatomical model: intrinsic and extrinsic factors

The patho-anatomical model is based upon structural and anatomical factors as the precipitator or source of pathology leading to a symptomatic presentation (Ludewig et al. 2013). Various elements have been proposed including intrinsic versus extrinsic factors.

Intrinsic factors are those which are concerned with the rotator cuff, including tendon morphology, biology and mechanical properties (Seitz et al. 2011). Conversely, extrinsic factors such as acromial architecture (primary impingement) and altered glenohumeral and scapular kinematics (secondary impingement) are seen as potentially precipitating or perpetuating pathology (Michener et al. 2003). Alongside these are scapular muscle and rotator cuff performance (Seitz et al. 2011). However there is acknowledgement that both intrinsic and extrinsic factors likely operate in unison as part of this multi-factorial pathology (Soslowsky et al. 2002; Seitz et al. 2011).

The model of compression and impingement is historically linked with sub-optimal acromion morphology leading to catching of tissues in the subacromial space for which excision of the lateral acromion and subacromial bursae can be advocated to provide sufficient subacromial clearance of the supraspinatus tendon (Neer 1972). However, altered scapular control could similarly give rise to tissue compression or insufficiency of the rotator cuff due to altered length-tension ratios (Pandey et al. 2016). Calcific deposits within the supraspinatus tendon may be considered to lead to impingement, whilst in their immature state calcific deposits can themselves be considered highly symptomatic (Micheroli et al. 2015). However, the non-definitive link between structural pathology and symptoms (Girish et al. 2011) complicates the drawing of such conclusions.

In addition, there has been increasing emphasis on movement system diagnoses rather than reliance upon patho-anatomic diagnoses (Lewis 2009; Ludewig et al. 2013). One of the drivers for this is the consistent reporting of poor specificity of orthopaedic tests commonly used to identify SIS/RCTendinopathy (Calis et al. 2000; Hegedus et al. 2008; Hegedus et al. 2012) which poses a challenge to the traditional diagnostic process of identifying SIS/RCTendinopathy patients.

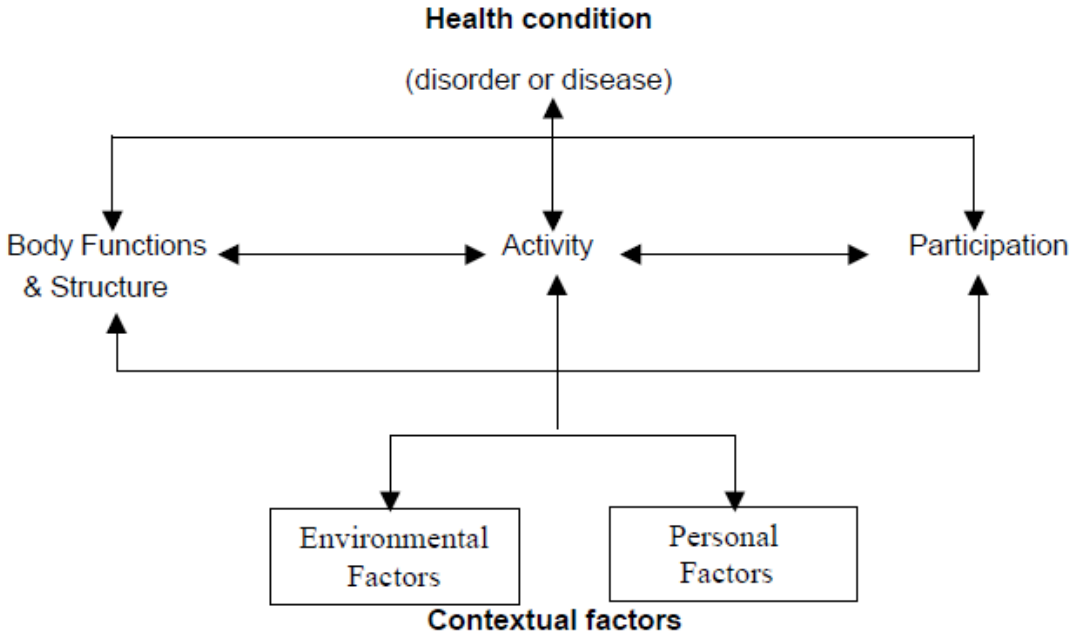
2.2.2 Psycho-social model

The psychosocial model considers psychological and work-related factors as being critical components in the precipitating or perpetuating of symptoms (Waddell 2004). Psychological factors such as anxiety, depression or distress can potentially have a widespread impact on willingness to use the shoulder in work or recreational activities and could impair treatment effectiveness (Vargas-Prada and Coggon 2015). Work-related elements can either predispose to or perpetuate chronic shoulder symptoms (van Rijn et al. 2010), including SIS/RCTendinopathy. As such the psychosocial model is recognised across musculoskeletal pathologies as highly relevant (Hill and Fritz 2011; Phyomaung et al. 2014) despite being potentially at odds with a strictly patho-anatomical model.

2.2.3 ICF classification

The International Classification of Functioning, Disability and Health (ICF) (WHO 2001) however provides a conceptual framework for functional measures within a bio-psycho-social context (Roe et al. 2013). As such it has the potential to classify according to a variety of inter-relationships (Lakke et al. 2009) as illustrated below:

Figure 2-1: ICF (WHO 2001)



These are further defined:

- Body Functions are physiological functions of body systems (including psychological functions).
- Body Structures are anatomical parts of the body such as organs, limbs and their components.
- Impairments are problems in body function or structure such as a significant deviation or loss.
- Activity is the execution of a task or action by an individual.
- Participation is involvement in a life situation.
- Activity Limitations are difficulties an individual may have in executing activities.
- Participation Restrictions are problems an individual may experience in involvement in life situations.
- Environmental Factors make up the physical, social and attitudinal environment in which people live and conduct their lives.

The ICF therefore provides a generic mechanism to capture a wide range of elements linked to shoulder pain, including SIS/RCTendinopathy (Roe et al. 2013).

2.3 Evidence regarding prognostic indicators of successful rehabilitation outcome in patients with SIS/RCTendinopathy

2.3.1 Search strategy

In order to identify research pertaining to the topic of 'prognostic indicators of successful rehabilitation outcome in patients with SIS/RCTendinopathy' the following databases were searched: AMED (Allied & Complementary Medicine), Biomed Central, CINAHL, The Cochrane Library, EMBASE, Google Scholar, Medline / PubMed, Medline via Ovid, PubMed, Scopus, Web of Science. For the purposes of this study, rehabilitation was defined as "non-invasive, multimodal physiotherapy".

The search terms and combinations were as follows: 1. exp Shoulder Impingement Syndrome/, 2. exp Shoulder/, 3. exp Shoulder Joint/, 4. exp Shoulder Pain/, 5. exp Rotator Cuff/, 6. exp Bursa, Synovial/, 7. exp Bursitis/, 8. exp Acromion/, 9. exp Tendinopathy/, 10. subacromial impingement.tw., 11. Shoulder Impingement.tw., 12. (Shoulder adj Pain).tw., 13. Rotator Cuff.tw., 14. (shoulder adj3 Bursitis).tw., 15. or/1-14, 16. exp Physical Therapy Specialty/, 17. exp Exercise Therapy/, 18. exp Physical Therapy Modalities/, 19. exp Rehabilitation/, 20. rehabilitat*.mp., 21. Physiotherap*.mp., 22. (physical adj therap*).mp., 23. or/16-22 24., cohort.mp. 25., prognostic factors.tw., 26. incidence.sh., 27. follow-up studies.sh., 28. prognos:.tw., 29. predict:.tw., 30. course:.tw., 31. or/24-30, 32. 15 and 23 and 31.

In parallel to the above main search, reference list screening and citation tracking was used as well as generalised web searching to identify potential 'grey' literature. The above searching was performed at the initiation of the project and the final search was undertaken on 3rd October 2016.

2.3.2 Identification of relevant studies

In relation to the topic of 'prognostic indicators of successful rehabilitation outcome in patients with SIS/RCTendinopathy', studies were considered relevant if they provided evidence relating to predicting outcome, i.e. prognosis. However identifying studies which could provide relevant evidence in relation to treatment and pathology was more difficult to definitively define.

In order to avoid an overly narrow review process but still ensure relevance, studies were included if they matched one of the following criteria:

(i) presented data relating to patients with SIS/RCTendinopathy who had been treated with non-invasive, multimodal physiotherapy

(ii) presented data relating to patients with non-specific shoulder pain who had been treated with non-invasive, multimodal physiotherapy

(iii) investigated patients with SIS/ RCTendinopathy or non-specific shoulder pain where treatments included non-invasive, multimodal physiotherapy AND outcome for those treated with non-invasive, multimodal physiotherapy could be identified from outcome where different interventions were applied.

Criteria (ii) and (iii) were included because there is controversy in the literature regarding the signs and symptoms of SIS/RCTendinopathy. Non-specific shoulder pain cohorts are highly likely to include patients with SIS/RCTendinopathy due to the high prevalence of SIS/RCTendinopathy in the general shoulder pain population (Ostor et al. 2005).

Table 2-1: Table of excluded studies and rationale

Rationale for exclusion	Studies excluded
Studies that were demonstrably concerned with shoulder pathologies that were discrete from SIS/RCTendinopathy; adhesive capsulitis	Binder et al. (1984), Shaffer et al. (1992), Mao et al. (1997), Griggs et al. (2000), Aydogan et al. (2003), Yang et al. (2008), Tanaka et al. (2010), Ando et al. (2013).
Studies that were demonstrably concerned with shoulder pathologies that were discrete from SIS/RCTendinopathy; atraumatic shoulder instability	Kuroda et al. (2001), Kim et al. (2004).
Studies that did not report segregated findings for the shoulder because these authors considered patients with pain in multiple, different body segments	Karels et al. (2007), Ryall et al. (2007).
Studies that did not present results segregated for those patients who were treated with non-invasive, multimodal physiotherapy, i.e. patients were managed with a variety of primary care approaches such as a 'wait-and-see policy', advice on avoidance of aggravating activities, oral medication, local injections of a steroid or anaesthetic or referral on for physiotherapy or surgery	Chard et al. (1988), Croft et al. (1996), van der Windt et al. (1996), Kuijpers et al. (2006a), Kuijpers et al. (2006b), van der Windt et al. (2007), Reilingh et al. (2008)
Studies where the intervention was steroid injection +/- non-invasive, multimodal physiotherapy and where results were not segregated for those patients who were treated solely with non-invasive, multimodal physiotherapy	Ginn and Cohen (2004), Zheng et al. (2005), Cummins et al. (2009), Ekeberg et al. (2010a), Cadogan et al. (2012), Contreras et al. (2013), Laslett et al. (2015)
Studies that were concerned with assessing outcome in occupational settings without consideration of if or how pathology was treated and/or where treatment details were	Macfarlane et al. (1998), Kaergaard and Andersen (2000), Viikari-Juntura et al. (2000), Miranda et al. (2001), Solomon et al. (2001),

not provided in sufficient detail to determine outcome in those treated with non-invasive, multimodal physiotherapy	Cassou et al. (2002), van Eijsden-Besseling et al. (2010).

* = examples: Kaergaard and Andersen (2000) explored the prognostic factors for continued musculoskeletal symptoms in sewing machine operators, but cases were not treated; instead differences were explored based on remaining in or leaving that type of work.
Miranda et al (2001) explored the prognostic factors for shoulder pain in Finnish forestry workers, but no details of treatment type was provided.

2.3.3 Relevant studies

Following the above processes, the studies by Bartolozzi et al. (1994), Morrison et al. (1997), Conroy and Hayes (1998), Kennedy et al. (2006), Deutscher et al. (2009), Ogon et al. (2009), Virta et al. (2009), Engebretsen et al. (2010), Hung et al. (2010), Mintken et al. (2010), Tyler et al. (2010), Sindhu et al. (2012), Kromer et al (2014) and Chester et al (2016) were identified. A critical assessment of the findings from these studies in relation to 'prognostic indicators of successful rehabilitation outcome in patients with SIS/RCTendinopathy' subsequently forms a large proportion of this literature review.

To complement the following critique of the literature, the characteristics of the studies are summarised in table form in appendix 2, mirroring the approach of Braun et al (2016).

2.3.4 Systematic approach to critical appraisal of the included studies

In order to provide a transparent, reproducible and systematic consideration of the strength of the evidence provided by each study, a recognised quality appraisal tool to assess study quality was used. Previous systematic reviews in this areas such as Chester et al (2013) have used bespoke checklists. However a formal checklist (PROBAST; Prediction model study Risk Of Bias Assessment Tool 2015) for assessing the risk of bias and assessing applicability was used in the most recent review in this area (Braun et al. 2016) and so this was also used to support the current literature review.

Details of how to complete the PROBAST assessment are provided in appendix II (PROBAST process), but in summary it provides a framework for assessing risk of bias and applicability of primary studies evaluating prognostic models. It comprises 5 domains that cover participant selection, predictors, outcome, sample size and participant flow and analysis (PROBAST 2015). The outcomes for steps 1-5 of the PROBAST process in relation to this literature review are provided in appendix II (PROBAST outcomes) and appendix II (Overall judgement on usability of the model).

As illustrated by appendix II (PROBAST Overall judgement), all studies were rated as being at high risk of bias and in two-thirds of studies there was high level concern regarding applicability to the research question. In addition, no studies were identified as being useable in the context of the intended context and target population. Although the systematic review by Braun et al. (2016) did not mirror exactly the studies included in the current literature review (the authors included studies of cuff tears and they did not exclude studies combining physiotherapy with other modalities), Braun et al. (2016) arrived at the same conclusion – namely that all studies were rated as being at high risk of bias and none were deemed usable in clinical practice. Consequently Braun et al. (2016) – and the previous systematic reviews in this area by van der Heijden (1999), Kuijpers et al. (2004) and Chester et al. (2013) – all undertook narrative analyses. Given the PROBAST findings in appendix II, this literature review will also adopt a narrative approach.

This critique of the literature has therefore been structured according to the key components of prognostic studies (Hemingway et al. 2013; Hingorani et al. 2013; Riley et al. 2013), as detailed below. This is to facilitate alignment of common elements across the chapters.

- the study design, including details of the pathology being investigated and where in the patient care pathway subjects were recruited; the sample size and nature of those who did and did not consent
- the potential prognostic factors and how they were measured
- the treatment intervention
- the outcome measures and how they were measured
- the analysis approach and the subsequent prognostic findings

2.4 Study design, pathology, patient care pathway and sample

2.4.1 Study design

Prognostic data can be derived from various study design formats. The studies by Conroy and Hayes (1998), Engebretsen et al. (2010) and Kromer et al. (2014) used secondary analysis of randomised control trialled (RCT) data. Whilst such data has value in the context of prognosis, RCTs can provide findings of limited transferability due to often high rates of non-recruitment to the study and the specified nature of the treatment (Moons et al. 2009a).

Bartolozzi et al. (1994), Morrison et al. (1997) and Sindhu et al. (2012) undertook retrospective studies but no mention was made of blinding, which raises the spectre of intentional or unintentional bias due to knowing the outcome prior to collecting and analysing the data. Mintken et al. (2010) performed prognostic analysis on data from their single-arm trial on a non-specific shoulder pain sample. Although this can be considered a cohort study design, the prescriptive intervention limits transferability to routine practice. Conversely the studies by Kennedy et al. (2006), Deutscher et al. (2009), Ogon et al. (2009), Virta et al. (2009), Hung et al. (2010), Tyler et al. (2010) and Chester et al. (2016) used prospective cohort designs which are ideal for maximising the prognostic rigour whilst controlling for potential bias (Moons et al. 2009a). In addition they allow for maximal clinical applicability because their inclusion criteria and treatment approaches are typically less narrowly defined compared to RCTs. Further aspects of the above studies will be considered before an overall judgment on quality is made.

2.4.2 Pathology (including inclusion and exclusion criteria)

The patient sample to be selected and specifically the pathology under investigation have clear ramifications for the design and interpretation of the research (Hemingway et al. 2013). As noted in the introduction chapter, pain experienced in the shoulder can be referred from the viscera and spinal structures (Linaker and Walker-Bone 2015) but it can be argued that this is discrete from pain originating from shoulder structures. Structures implicated in pain originating from the shoulder are the peripheral receptors (particularly within the highly innervated subacromial bursae), peripheral pain processing, the spinal cord and the brain (Dean et al. 2013)). Pain originating from shoulder structures can encompass non-specific shoulder pain alongside attempts to categorise the nature of the pain. Numerous such categories of local shoulder pain are described, comprising instability, adhesive capsulitis, acromioclavicular joint (ACJ) pain, osteoarthritis, rotator cuff disease and impingement type symptoms (Schellingerhout et al. 2008).

Gross shoulder instability can be viewed as a relatively discrete clinical and aetiological condition whereby structural or muscle patterning alterations can give rise to observable instability at the glenohumeral joint (GHJ) (Hayes et al. 2002). Adhesive capsulitis is typically characterised by stiffness of the shoulder complex, particularly the GHJ. Although the aetiology of adhesive capsulitis remains uncertain (Linaker and Walker-Bone 2015), the clinical presentation of adhesive capsulitis with its characteristic freeze and thaw phases is reported. ACJ pain whereby symptoms are localised to the ACJ and replicated upon approximation of the joint surfaces (Armstrong 2014) can also be seen as a relatively discrete condition. Osteoarthritis of the GHJ is typically a radiological diagnosis

(Parsons et al. 2004) and whilst it accounts for only a limited percentage of shoulder pain patients, it can be readily classified.

However, there is arguably greater uncertainty regarding SIS/RCTendinopathy type symptoms (Walker-Bone et al. 2003). In SIS/RCTendinopathy the classical concept of impingement describes the compression of tissues within the subacromial space and the local generation of painful stimuli (Neer 1983). These provide the rationale for orthopaedic symptom reproduction tests such as the Hawkins-Kennedy test, Neer's sign and the Empty Can test (Alqunaee et al. 2012). These tests are often described as compressing the irritated, tendinopathic or torn tendon tissue that lies in this space. However, the identification of high levels of nociceptors in the bursae lining this region provides a rationale for bursal tissue as being a primary source of symptom production (Lewis 2009). Furthermore, long-term exposure to pain can arguably give rise to symptoms that have limited if any local source but instead are centrally generated (Kooijman et al. 2015).

Whilst the importance of the rotator cuff from a neuromotor control perspective is widely accepted, its role in pathology is more controversial (Seitz et al. 2011). Sonographically the presence of hypo and hypertrophy of the rotator cuff tendon may be theorised to be associated with under or over-use, respectively. Tendinopathic changes such as decreased echogenicity and internal architecture heterogeneity can lead to a diagnosis of rotator cuff tendinopathy (Smith et al 2015). However, the relative relevance of calcific deposits (Gosens and Hofstee 2009) and associated findings such as bursal thickening and increased fluid are areas of controversy. The theorised linking of such findings with mechanical 'impingement' is often proposed, with this historically being linked to acromion morphology (Pandey et al. 2016). The presence of rotator cuff tears could be viewed as a progression of such rotator cuff disease (Lewis 2010) although the mechanisms by which this might occur are debatable, as are the symptomatic relevance of such tears. In light of the uncertainties as to the aetiology and patho-anatomical processes involved in SIS/RCTendinopathy, defining this pathology via inclusion and exclusion criteria is therefore an area of controversy.

For the studies by Kennedy et al. (2006), Deutscher et al. (2009), Mintken et al. (2010) and Chester et al. (2016) which were concerned with non-specific shoulder patient populations then identification was typically via exclusion of visceral, systemic or spinal sources of symptoms. This leaves local shoulder reproduced symptoms, which is a relatively easily defined population. However in terms of the likely predictors of outcome and particularly the clinical application of the findings, a generalist approach can be argued to be of limited utility. This is further compounded by the

potential that quite different treatment approaches may be applied to distinctly different pathologies, e.g. excessive stiffness (e.g. adhesive capsulitis) versus excessive movement (e.g. gross instability). This additional heterogeneity within the data could theoretically make it less likely to identify clinical or pathology-related factors as predictors. Nonetheless, these studies would likely include a substantial number of SIS/RCTendinopathy patients due to their consideration of non-specific shoulder pain populations (van der Windt et al. 1995; Ostor et al. 2005). As such they were included in the review so as to avoid omitting potentially useful evidence.

An alternative approach was taken by Sindhu et al. (2012) who used the diagnosis codes based on the International Classification of Diseases, Ninth Revision (ICD-9) to identify different categories of non-specific shoulder pain, of which the 'muscle, tendon and soft tissues disorders' were deemed most comparable to the pathologies under consideration in this study. However the limited inter-rater agreement for the ICD-9 system (Dixon et al. 1998) raises questions as to the merit of such an approach.

Conversely where studies specified particular pathologies these included the labels of subacromial impingement (Morrison et al. 1997; Virta et al. 2009; Hung et al. 2010), subacromial shoulder pain (Engebretsen et al. 2010), shoulder impingement syndrome (Conroy and Hayes 1998), subacromial pain syndrome (Kromer et al. 2014), rotator cuff disease (Bartolozzi et al. 1994) and calcific tendonitis (Ogon et al. 2009). This reflects uncertainty as to the nature of the pathology and associated variation in the terminology used.

However when looking at the actual inclusion and exclusion criteria there were areas of commonality. Typically, inclusion criteria involved orthopaedic tests results with Morrison et al. (1997) requiring a positive Neer's sign and Engebretsen et al. (2010) requiring a positive Hawkins-Kennedy test. Other studies such as Kromer et al. (2014) required presence of 1 positive test from a battery of tests (Neer's sign, Hawkins-Kennedy test, or painful arc with active abduction or flexion) while Bartolozzi et al. (1994) required a positive Neer's sign, Hawkins-Kennedy test and a painful arc in those who did not have imaging. Hung et al. (2010) also used a battery of tests (Neer's sign, Hawkins-Kennedy test, painful arc, pain with palpation of rotator cuff tendons and pain with active shoulder elevation) but required 3 or more to be positive for inclusion in the study.

However the sensitivity and specificity of orthopaedic tests has been consistently identified as being limited (Calis et al. (2000); Alqunaee et al. (2012)). Such findings emphasise that even using a battery

of tests, identifying a patient population where orthopaedic test results are an inclusion criterion is arguably flawed. As a consequence, identification of a sample based on clinical history and clinical examination (but not on orthopaedic test results) can be considered of greater clinical and conceptual relevance (Kromer et al. 2010).

In this respect, clinical presentation suggestive of the target pathology was used as an inclusion criterion by Bartolozzi et al. (1994), Morrison et al. (1997) and Ogon et al. (2009). This included Engebretsen et al. (2010) and Kromer et al. (2014) requiring dysfunction or pain on shoulder abduction and pain on isometric tests of abduction, external and internal rotation and Conroy and Hayes (1998) requiring pain in the supero-lateral shoulder region.

From an exclusion perspective, elimination of other discrete pathologies such as instability, adhesive capsulitis, acromio-clavicular joint pain and GHJ arthritis was used by Conroy and Hayes (1998), Engebretsen et al. (2010) and Kromer et al. (2014); Bartolozzi et al. (1994), Morrison et al. (1997), Conroy and Hayes (1998) and Kromer et al. (2014); Morrison et al. (1997), Conroy and Hayes (1998) and Engebretsen et al. (2010); Bartolozzi et al. (1994), Conroy and Hayes (1998) and Engebretsen et al. (2010), respectively. As such, exclusion of other discrete pathologies is a commonly used mechanism.

The above section presents the variety of approaches taken for identifying a patient sample, many of which overlap to a greater or lesser extent. However this heterogeneity highlights the challenges of prognostic research in SIS/RCTendinopathy patients, in that uncertainty persists as to the relevance or otherwise of different pathological characteristics. As such, these factors highlight potential challenges in identifying predictors of outcome.

2.4.3 Patient care pathway and clinical setting

This thesis is focused on physiotherapy, but it is of note that various treatment approaches for patients with SIS/RCTendinopathy may be offered and that physiotherapy treatment can occur at one or more of a variety of points within the care pathway.

Typically, most patients will initially present in primary care, usually to their GP (Ostor et al. 2005; Dorrestijn et al. 2011). Treatment options here include advice, rest, oral medication (analgesia and/or anti-inflammatories), injection of the shoulder (typically with a combination of analgesia and anti-inflammatories) (Gruson et al. 2008) and referral on (Mitchell et al. 2005).

Other specialities involved include Orthopaedic surgery and Rheumatology where available treatment modalities include surgery such as subacromial decompression with or without rotator cuff repair (Neer 1972; Henkus et al. 2009; Butt et al. 2015; Carr et al. 2015), shoulder injections (image guided or unguided) (Dogu et al. 2012; Min et al. 2013; Haghighat et al. 2016; Messina et al. 2016) and referral on.

As noted above, where patients are referred for rehabilitation, typically in the form of physiotherapy in the NHS, this can be directly from their GP or from other care providers based in secondary care (Dorrestijn et al. 2011; Kooijman et al. 2013). However the care pathway for shoulder pain and SIS/RCTendinopathy is typically poorly defined and may be more influenced by clinical availability, patient preference and symptom persistence, rather than a definitive, optimal patient management pathway.

In the studies under consideration, patients were recruited from a variety of setting. Studies such as Kennedy et al. (2006) recruited patients who were beginning treatment at a large number of different physiotherapy practices. Mintken et al. (2010) also recruited from outpatient physiotherapy clinics whilst Hung et al. (2010) and Engebretsen et al. (2010) recruited from orthopaedic clinics and university hospital outpatient departments; respectively. Others such as Kromer et al. (2014) recruited their sample from outpatient physiotherapy clinics, GPs and orthopaedic surgeons. The above details are evidence of patient recruitment heterogeneity in relation to care pathway.

In addition to this, Virta et al. (2009) recruited patients who were on a waiting list for orthopaedic surgery but had been referred for physiotherapy prior to surgery. Similarly Ogon et al. (2009) only included patients who had been referred following failed conservative treatment, which comprised physiotherapy, manual therapy, electrotherapy, oral analgesics or NSAIDs and up to three subacromial corticosteroid injections. It could therefore be argued that both studies were concerned with patients at a relatively late stage in the care pathway and where patient expectations of physiotherapy may have been low.

Such methodological elements have obvious implications in terms of likelihood of success as well as transferability of the findings. In order to optimise the usefulness of study findings, clearly defined

care pathways would increase the likelihood of meaningful results and also the ability to apply any findings in a more targeted way for clinical implementation.

2.4.4 Sample size and nature of those who did and did not consent

As in other types of research, sample size is an important consideration particularly in relation to the statistical power to accurately detect factors that are or are not predictive of outcome. However the challenges around sample size calculations are manifold (Moons et al. 2009a; Riley et al. 2013) and this is reflected by the lack of sample size justification in the majority of prognostic studies in this area.

Typically where sample size calculations were presented these related to determining the sample required for differences to be identified in RCTs, such as by Engebretsen et al. (2010) and Kromer et al. (2014). However with cohort studies a commonly used approach is the guide of 10 cases per prognostic factor (Peduzzi et al. 1996). One challenge here is that prior to the onset of a study it is not possible to identify the number of variables that will be retained for entering into the prognostic model (depending on how the model is populated). This might help to explain the predominance of seemingly pragmatic sample sizes. As such these are likely influenced by the use of retrospective electronic databases which facilitate collection of large samples (e.g. Deutscher et al. 2009 with n=5,000) through to complex three-dimensional motion analysis requiring lengthy data collection phases where smaller sample sizes are seen (e.g. Hung et al. 2010 with n=33).

An alternative, pragmatic approach is to use a combination of statistical and conceptual methods to align the number of potential predictor variables with the size of the sample. This has the advantage of reducing the risk of overfitting, whereby the inclusion of too many potential predictors for the sample size makes it unlikely that when model validation is undertaken in a new sample that the original results will be reproduced (Kent et al. 2010). Indeed this has been cited as one of the greatest threats to the robustness of prognostic research, particularly in smaller sample sizes (Babyak 2004).

Alongside the sample size, the characteristics of those who did and did not consent should be considered, so as to avoid potential selection bias. A comparison of characteristics between groups was not performed by the majority of studies and so any potential bias is unknown. However the studies by Kennedy et al. (2006) and Chester et al. (2016) did undertake statistical analysis and identified no difference in gender or age between the groups.

2.5 Potential prognostic factors and how they were measured

The potential prognostic factors selected and the methods by which they are derived, processed and utilised in prognostic modelling have substantial implications for study findings and interpretation (Hemingway et al. 2009). Typically, potential predictors are assessed at baseline such that the subsequent results can inform how baseline characteristics can influence outcome (Moons et al. 2009b). The variables to be considered can be across a spectrum of elements either identified as being predictive in previous studies or conceptually associated with the condition, management or outcome.

In the case of the identified prognostic studies for review, a wide range and number of variables have been considered. In order to allow for meaningful comparison within categories, the potential prognostic factors considered in the relevant published literature have been collated according to (A) demographics, (B) clinical history, (C) patient reported measures, (D) clinical measures and (E) structural pathology.

In the following section the strengths and weaknesses of the approaches used in previous studies will be considered. Later in this chapter the actual variables which were and were not identified as being predictive in individual studies will be presented.

2.5.1 A. Demographics

Age and gender were the demographic variables most commonly explored as prognostic factors by studies in this topic area. Indeed all studies that explored demographic factors as potential predictors considered age, while only Conroy and Hayes (1998) and Hung et al. (2010) omitted gender from their analysis. For the later study this was due to their use of a male only sample. Kromer et al. (2014) recognised a potential role for these variables but chose to control for both age and gender when constructing their prognostic model.

Along with the potential association of increasing age with poorer tissue healing, age could be rationalised as a predictor due to theorised associations with pre-existing activity levels, greater functional reserves and enhanced ability to modify lifestyle amongst the young (Cecchi et al. 2014). Analysis of age as a continuous variable facilitates retention of maximal statistical power and is advocated for robust prognostic analysis (Hingorani et al. 2013). The decision by studies such as

Bartolozzi et al. (1994), Morrison et al. (1997) and Kennedy et al. (2006) to convert their continuous variable of age into an ordinal variable is therefore controversial due to a subsequent reduction in statistical power.

A potential role for gender as a predictor for outcome comes from evidence of differences between the sexes for pain sensitivity (Horn et al. 2014) including in the shoulder (Kindler et al. 2011). The evidence generally points to a higher incidence of pain conditions and sensitivity in women compared to men, although conflicting evidence predominates in response to differences between the genders for pain treatment (Fillingim et al. 2009). Evidence of a gender effect in the low back pain prognostic literature (George et al. 2006) supports the potential cross-condition and generic relevance of this variable.

A less commonly explored generic demographic variable was smoker status which was only considered by those studies that investigated general shoulder pathology (Deutscher et al. 2009; Chester et al. 2016). The potential impact on tissue healing provides a rationale for considering this variable although uncertainty regarding the direction of cause and effect with regards to smoking and pain confounds establishing a consensus in this area (Ditre et al. 2011).

The study by Chester et al. (2016) was the only one to consider the variables of body mass index (BMI) and social deprivation. Associations between elevated BMI and both a higher incidence of and poorer recovery from musculoskeletal disorders (Viester et al. 2013) provides a rationale for considering this variable. Socioeconomic status has been identified as impacting upon morbidity arising from musculoskeletal disorders (Brekke et al. 2002), thereby providing a rationale for exploring it as a predictor. However the use by Chester et al. (2016) of the index of multiple deprivation as a representative variable can be considered of limited accuracy as it is postcode based rather than individualised to the patient. Another representative variable (educational attainment) was only explored by Engebretsen et al. (2010) who defined this as whether the patient had attended college or university. An association of educational attainment with musculoskeletal pain and disability (Katz 2006) provides a rationale for considering this variable.

One or more of work, recreational or functional task related variables were included by 8 studies, with occupation and/or work status recorded by Bartolozzi et al. (1994), Kennedy et al. (2006), Ogon et al. (2009), Engebretsen et al. (2010), Mintken et al. (2010) and Chester et al. (2016). Physical activity levels were recorded by Deutscher et al. (2009) and Chester et al. (2016) whilst functional

tasks were considered by Engebretsen et al. (2010) and Chester et al. (2016). However studies such as Ogon et al. (2009) provided no details of how such variables were collected or defined. Conversely Chester et al. (2016) used a robust measure to quantify activity levels, namely the Godin leisure time exercise questionnaire (Godin and Shephard 1985). Deutscher et al. (2009) on the other hand simply recorded the number of times per week patients said they were active, thereby providing less robust data.

There was also inconsistency with how occupation and/or work status was derived. Kennedy et al. (2006) looked at whether the patient was unable to work due to their shoulder pain whilst Engebretsen et al. (2010) and Chester et al. (2016) looked at the type of work, the intensity and likely symptom aggravating nature of their work or recreational activities. Such inconsistencies in the methods of data collection inevitably present a potential barrier to deriving a consensus from the literature in this area.

It can therefore be concluded that a wide variety of demographic related variables have been considered in previous prognostic studies. There is merit is collecting such data in order to adequately describe the sample. However a judicious choice should be made as to the demographic variable or variables to be entered into a subsequent prognostic model so as to avoid potential over-fitting.

2.5.2 B. Clinical history

Symptom duration was the most commonly investigated 'current condition related variable' and can be justified by evidence of more established conditions being more stubborn to treatment (Collins et al. 2010). In addition the inter-relationship of chronicity, depression and pain levels (Sullivan et al. 2008); and identification of symptom duration as a predictor across various musculoskeletal disorders (Rihn et al. 2011; McClinton et al. 2015) provide further rationale for considering it.

However the handling of this patient reported variable was a source of less consistency, including dichotomising to acute versus chronic (as per Deutscher et al. 2009), the use of categories of time and retention of it as a continuous variable. In justifying the use of soft-tissue healing timescales, Kennedy et al. (2006) converted their duration data into less than 4 weeks, 4 to 12 weeks and longer than 12 weeks. However controversy regarding tissue-healing timescales (Broughton et al. 2006) and the complex mechanisms underpinning SIS/RCTendinopathy symptoms make such arbitrary timescales arguably of limited utility. Conversely, retaining such data as a continuous variable has

the advantage of preserving optimal statistical power (Hingorani et al. 2013). However absolute numerical data is potentially subject to recall error, particularly where onset is not attributed to a single, discrete event. As such, the use of an ordinal category as per Engebretsen et al. (2010) and Virta et al. (2009) provides a reasonable compromise.

Side-related symptoms were also commonly investigated, with five studies reporting data in relation to arm dominance. Conversely Chester et al. (2016) explored bi-laterality and the side of the symptoms relative to the non-involved side, making it difficult to compare results directly between studies. A similar number of publications (n=5) explored the reason for symptom onset although as this was patient reported it is inherently subject to potential recall bias or inaccuracy. Recurrence of the shoulder symptoms was explored as a potential predictor by four studies although again as patient reported data this comes with the caveats previously mentioned. Associated symptoms of neck pain and paraesthesia in the arm were explored by Engebretsen et al. (2010) and Chester et al. (2016); respectively. The Engebretsen et al. (2010) data was patient reported, again raising the spectre of recall error whilst Chester et al. (2016) failed to clarify whether paraesthesia was determined by the patient or clinician.

In regard to treatment related variables, the variable of previously receiving treatment (including physiotherapy) might indicate that treatment again will be successful or unsuccessful, depending on the previous therapeutic response. As such, Engebretsen et al. (2010), Mintken et al. (2010) and Chester et al. (2016) explored this variable. Taking medication for pain or psychological state might also reflect either a more complex clinical presentation or conversely that other condition related symptoms are being actively managed (Ndlovu et al. 2014). As such, Kennedy et al. (2006), Deutscher et al. (2009), Engebretsen et al. (2010), Mintken et al. (2010) and Chester et al. (2016) all considered this variable. Finally, the presence of co-morbidities might conceivably represent a further level of clinical complexity and as such were explored by five studies.

As with the demographic variables, it can be concluded that a wide variety of clinical history related variables have been considered in previous prognostic studies. Recognising this and in relation to the use of this variable type in a prognostic study, a judicious choice should be made as to the variable or variables to be entered into a prognostic model so as to avoid potential over-fitting. Notwithstanding this, recording of clinical history related variables are also of value in adequately describing the sample.

2.5.3 C1. Patient reported measures: Pain

Alongside functional restrictions, pain is typically the major symptom that patients with SIS/RCTendinopathy report. Consequently the management and/or resolution of pain are major aims in the healthcare management of SIS/RCTendinopathy (Baring et al. 2007). In the context of prognosis, substantial pain levels at baseline could be theorised to be associated with worse outcome due to high pain levels potentially representing severity of the condition along with the potential impact of high pain levels on other aspects such as function, strength and wellbeing (Kooijman et al. 2015).

The studies by Kennedy et al. (2006), Engebretsen et al. (2010), Mintken et al. (2010), Kromer et al. (2014) and Chester et al. (2016) considered pain as a potential predictor and all measured pain levels via patient self-reported scores. However the complexities of how painful stimuli are generated, transmitted and received are only partly established (Chien et al. 2013). Furthermore, pain is recognised as a largely subjective experience which can be impacted upon by aspects such as personal belief mechanisms, societal conditioning and mental health, all of which add to the complexity of the measurement and treatment of pain (Hjermstad et al. 2011). As such, measuring pain is a challenging area (Ferreira-Valente et al. 2011).

One commonly used clinical approach is to ask patients to rate their symptoms using a scale which can be Likert, numerical rating or visual analogue in nature (Hjermstad et al. 2011), such as the Visual Analogue Scale (VAS). These have the advantage of being quick to perform but can be subject to recall bias and subjective interpretation. Whilst recognising these caveats, the VAS has been widely tested for acceptability (Joyce et al. 1975), reliability (Ferraz et al. 1990), construct validity when compared to verbal descriptive and numeric rating scales (Downie et al. 1978). A further advantage of this approach is that minimum clinically important difference data has been reported for patients with rotator cuff disease managed non-operatively (Tashjian et al. 2009) making it a relevant approach to use.

Engebretsen et al. (2010) asked patients to rate their pain according to the levels ‘at rest’ and ‘during activity’ in the previous week. However they failed to stipulate whether the ‘during activity’ rating given by the patient was to refer to pain levels during generalised shoulder activity or in relation to a pain reproducing shoulder activity. Such a variation in potential interpretations makes the ‘during activity’ pain data in this study of questionable value. Chester et al. (2016) recorded only pain levels at rest, although whether this was momentary or over a time period was not stipulated.

Conversely Kennedy et al. (2006) and Kromer et al. (2014) made no attempt to stipulate what aspect of rest or activity their patients should report their pain levels in relation to, asking instead that subjects simply gave the rating that best described their average pain level. Such an approach is non-specific considering the potential for substantial symptom variability.

The use by Kennedy et al. (2006), Engebretsen et al. (2010) and Kromer et al. (2014) of 1 week as the period over which patients were asked to rate their pain has the advantage of potentially accommodating day to day variation in pain levels. However the use of such a time period might introduce recall error because accurately recalling pain levels across a 7-day period could be prone to inaccuracy. Furthermore, in the case of Engebretsen et al. (2010) it also raises the question of whether patients reported their lowest, greatest or average pain levels across the time period.

The data processing and presentation of the pain variables for potential inclusion in the final prognostic model also varied across different studies. Most studies simply reported their individual pain variables whilst Mintken et al. (2010) reported only the mean of their three variables (current pain along with their worst and least amount of pain in the previous 24 hours). This approach has the advantage of accommodating likely variation in pain across time periods and activities, and therefore arguably being more globally representative.

It can be concluded that whilst there is consistency in the choice of patient self-reporting as the mechanism for measuring pain, the specifics of how this has been carried out varies widely across previously published prognostic studies. In relation to the measuring of this variable in a prognostic study, the accurate stipulating of the aspect of pain (e.g. lowest, greatest, average) and what it is in relation to (e.g. rest, activity, a test movement, etc.) is advocated. Similarly, recognition of potential recall error when considering the time period over which to rate pain and the segregating of rest, activity and night pain is of merit. The use of a combination or mean score as a mechanism for accommodating the wide variation across such segregations also has merit.

2.5.4 C2. Patient reported measures: Psychological symptoms

A patient's state of mind has the potential to impact upon their perception of subjective experiences such as pain and their self-identification of impairment levels. It can also influence the willingness of patients to use their shoulder and their willingness to participate in work, recreational activities and even treatment (Vargas-Prada and Coggon 2015). As such, psychological symptoms such as elevated

levels of anxiety, depression, somatisation or distress can potentially have a widespread impact (Vranceanu et al. 2009) and where elevated at baseline could be theorised to be associated with worse outcome. In the context of musculoskeletal pain, psychological symptoms might be pre-existing or might be precipitated or worsened by ongoing pain and impairment.

The studies by Kennedy et al. (2006), Engebretsen et al. (2010), Mintken et al. (2010) and Chester et al. (2016) used outcome measures of elements which are linked to state of mind but are not themselves psychological symptoms. Specifically these were self-efficacy for pain (Engebretsen et al. 2010; Chester et al. 2016), expectation (Kennedy et al. 2006; Mintken et al. 2010; Chester et al. 2016) and perceived general health status (Engebretsen et al. 2010).

Engebretsen et al. (2010) reported perceived self-efficacy data for pain comprising a mean score from four items rated from 1 (easy) to 7 (impossible) concerning self-judgments of performance capability. However they provided no further detail of the tool other than referring to the paper by (Lorig et al. 1989) who described a tool validated for patients with arthritis comprising 20 questions across 3 subscales, each scored on a scale from 10 (very uncertain) to 100 (very certain). The uncertainty as to the nature of the tool used by Engebretsen et al. (2010) and how it was derived means that such data cannot be relied upon. Conversely Chester et al. (2016) cited the thoroughly evaluated version reported by Nicholas (2007).

Kennedy et al. (2006) asked patients to rate their expectations for recovery under the domains of (i) how much they would recover and (ii) how quickly they would return to usual activities. The inclusion of patient expectation has merit in that expectation could be theorised to influence the level of participation in non-invasive, multimodal physiotherapy and the potential for high or low expectation of outcome to be a self-fulfilling prophecy (Vargas-Prada and Coggon 2015). However Kennedy et al. (2006) provided no evidence as to the specific question wording used nor the psychometrics of their tool and so it is difficult to gauge the value of the evidence provided in relation to this variable. Similarly Chester et al. (2016) stated their wording “how much do you expect your shoulder problem to change as a result of physiotherapy treatment?” but provided no reference for it. Mintken et al. (2010) also included expectation although they provided no further details beyond “The historical examination included ….. expectations for treatment …..” (page 29) and so it is not possible to gauge how this was collected or the subsequent nature of the data relating to this potential prognostic variable.

The study by Engebretsen et al. (2010) used an outcome measure that looked at other global elements which arguably incorporate state of mind. Specifically they evaluated general health status by using the EuroQol-visual analogue scale (EQ-VAS), a self-rated score for health on a vertical, visual analogue scale where the endpoints are labelled 'Best imaginable health state' and 'Worst imaginable health state' with higher scores indicating better perceived health (Brooks 1996). This internationally used outcome score has been demonstrated to be valid and reliable, including in musculoskeletal pain (Hurst et al. 1997). However it is an inherently generalist measure of self-perceived general health and so provides limited insight into or quantification of psychological symptoms. This is relevant because poor mental health can have a wide-reaching impact on, and association with, musculoskeletal pain and disability (Bair et al. 2003; Pincus et al. 2002; Phyomaung et al. 2014).

The more direct measures of state of mind or psychological symptoms included in the studies by Kennedy et al. (2006), Engebretsen et al. (2010), Mintken et al. (2010), Sindhu et al. (2012), Kromer et al. (2014) and Chester et al. (2016) can therefore be considered of greater relevance. Psychological symptoms in relation to fear of pain and consequent avoidance of physical activity because of fear were measured by Mintken et al. (2010), Kromer et al. (2014) and Sindhu et al. (2012) who used the Fear Avoidance Beliefs Questionnaire (FABQ) (Waddell et al. 1993). However there was wide variation in how this tool was described and applied.

The FABQ examines high pain avoidance behaviour and generates 'general, physical activity and work' subscales (representing beliefs about how work and physical activity impact upon their pain) from which a total score is calculated (Waddell et al. 1993). Its potential relevance as a predictor of outcome in patients with low back pain receiving physiotherapy has been previously demonstrated by George et al. (2008). However Mintken et al. (2010) and Kromer et al. (2014) modified the FABQ by changing the word "back" to "shoulder" on the questionnaire. It is unclear whether the validity of the questionnaire in a different anatomical region has been established and so the robustness of the data for this potential prognostic variable is uncertain. Furthermore Kromer et al. (2014) only chose to use a component of it, namely the physical activity subscale.

Sindhu et al. (2012) took this further and chose to only use a single item from the physical activities scale of the FABQ, namely "I should not do physical activities that (might) make my pain worse". The use of a single item was justified by the study by Hart et al. (2009) whose data supported this item as accurately identifying subjects with elevated levels of fear. This was appropriate for the unusual

study design used by Sindhu et al. (2012) as the authors were interested in exploring fear avoidance as a potential predictor of functional outcome across different patient groups but whilst controlling for other potential predictors. However the use of a dichotomised, single item greatly limits the breadth and sensitivity of the original FABQ, thereby providing very limited consideration of psychological symptoms of subjects in this study.

Alongside the FABQ, Mintken et al. (2010) also used an 11-item, shortened version of the Tampa Scale for Kinesiophobia (TSK) which is similar to the FABQ in that it assesses fear of both movement and pain. It has been previously used as a potential predictor of outcome in shoulder pain patients (Bot et al. 2005) and high test-retest reliability, good internal consistency and concurrent validity has been reported (Woby et al. 2005). Along with the FABQ, the relevance of the TSK to a patient population undergoing physiotherapy for shoulder pain is evident as it assesses fear of movement and pain. However it could be argued that this is also the major weakness of measures such as the FABQ and TSK in that complex, multi-faceted aspects of psychological symptoms will potentially be missed by such specific screening tools.

An alternative approach to quantifying psychological symptoms was used in the study by Kennedy et al. (2006) as the authors used subsections of another outcome measure to derive state of mind measures. Specifically Kennedy et al. (2006) used the Mental Component Score (MCS) from the 36-Item Short-Form Health Survey (SF-36). The SF-36 is a widely used health-related outcome measure and reliability of the component scales and summary measures have been demonstrated (McHorney et al. 1994), including the validity of the MCS (Ware et al. 1995). However the MCS comprises a broad range of subsections from the SF-36, i.e. Vitality, Mental Health, Emotional Role, and Social Functioning and so the MCS by its very nature can be considered a very general assessment of a respondents' state of mind but with limited specificity in relation to psychological symptoms *per se*.

Chester et al. (2016) reported data pertaining to anxiety and depression over the last 7 days but failed to detail how their rating of 'No, moderately and extremely' were determined. Considering the complex and sensitive nature of such elements the limited ability for a single question to capture this, raises doubts as to the value of their data. Conversely Kromer et al. (2014) took a more robust approach in using a multi-faceted, formal measure, namely the Pain Catastrophizing Scale (Sullivan et al. 1995).

Engebretsen et al. (2010) included a specific measure of psychological symptoms, which was the 25-items Hopkins Symptoms Checklist to assess emotional distress with a high score indicating more distress. However there is a lack of clarity as to which elements the measurement tool actually measured because they referenced Derogatis et al. (1974) which is the original 58-item tool for scoring somatisation, obsessive-compulsive, interpersonal sensitivity, anxiety and depression. This contrasts with the 25-item scale which comprises 10 anxiety items, 13 depression items and 2 somatisation (Winokur et al. 1984).

The inclusion of an outcome measure specific to psychological symptoms is a notable strength of the study by Engebretsen et al. (2010). However distress is only one of the mental health symptoms that patients may present with, alongside depression, anxiety and somatisation (den Boeft et al. 2016). The use of a measure of psychological symptoms such as via the four-dimensional symptom questionnaire (4DSQ) therefore provides a mechanism for measuring the broad range of psychological symptoms. The 4DSQ is specifically designed to measure psychological symptoms and sub-categorise the nature of them according to levels of anxiety, depression, distress and somatisation. Cut-off points have been published for the 4DSQ to guide the clinical interpretation of levels of psychological symptoms (Terluin et al. 2006). Collapsing of the moderately and strongly elevated categories into a single score to illustrate prevalence has also been published (Koorevaar et al. 2016). Alongside evidence of its validity (Tebbe et al. 2013), the appropriateness of its use in orthopaedic shoulder patients has also been demonstrated (Koorevaar et al. 2016).

It can be concluded that the inclusion of a specific assessment of psychological symptoms as a potential prognostic factor has merit so that the potential to impact upon prognosis and outcome can be explored. In relation to study design, it is proposed that diffuse aspects of state of mind such as perception, expectation and general well-being may be relevant but can be considered to be of limited specificity in relation to psychological symptoms. Movement or pain specific elements can be considered relevant to MSK conditions but limited in their ability to capture the full range of psychological symptoms that patients may present with. The use therefore of a measurement tool where the full breadth of psychological symptoms can be assessed is therefore advocated.

2.5.5 C3. Patient reported measures: Function / Disability

The level of shoulder function that a patient demonstrates is a composite of their physical ability (i.e. shoulder strength and range of movement) and how much they are able or willing to use that physical ability due to pain, perception and state of mind (Roe et al. 2013). Function and disability

are therefore wide ranging aspects of the patient experience and consequently encompass many of the elements that prompt patients to seek treatment for their shoulder (Payne and Michener 2014). As such, the optimising of function and addressing of disability are major aims in the healthcare management of shoulder pain, including SIS/RCTendinopathy.

In the context of prognosis, substantial impairment of function at baseline could be theorised to be associated with worse outcome due to the need to address substantial impairment, making for a more challenging rehabilitative process (Schmidt et al. 2014).

Many shoulder prognostic studies have utilised function as their primary outcome measure, i.e. the mechanism by which improvement or otherwise is determined. Where studies have considered function in that capacity, these will be considered later in the literature review. However the prognostic studies by Bartolozzi et al. (1994), Kennedy et al. (2006), Deutscher et al. (2009), Engebretsen et al. (2010), Hung et al. (2010), Mintken et al. (2010), Kromer et al. (2014) and Chester et al. (2016) included assessment of function as a potential prognostic factor and so these will be considered here.

The study by Kennedy et al. (2006) used a single item, global rating of disability score. The authors provided no details beyond stating that the variable was 'Patient global rating of shoulder injury/problem' scored as 'very mild, mild, moderate, severe or serious, very severe or serious' (page 1,018) and so the nature, reliability and validity of this variable cannot be determined. Kennedy et al. (2006) also used a derivation of another outcome measure to derive function measures. Specifically they used the Physical Component Score (PCS) from the 36-Item SF-36. It was noted in the section concerning psychological symptoms that the SF-36, its derivations and components have been widely tested. However the use of generic, quality of life scores means that whilst function related symptoms *per se* can be assessed, the measurement approach provides very limited shoulder-specific function information.

In contrast to a generic measure, Bartolozzi et al. (1994) used a shoulder specific functional outcome score, the University of California Los Angeles (UCLA) score. This 5-section questionnaire considers pain, function, active range of movement (ROM), strength and patient satisfaction (Amstutz et al. 1981) but Bartolozzi et al. (1994) chose to only use the function sub-section, which has not been separately tested for reliability or validity.

Conversely Hung et al. (2010) used the Flexilevel Scale of Shoulder Function (FLEXSF) whereby respondents answer a single question that grossly classifies their level of function as low, medium or high; respondents are then directed to only the items that target their level of function. Evidence of good reliability, validity, and responsiveness to clinical change has been reported (Cook et al. 2003). Similarly Deutscher et al. (2009) used the shoulder Computerized Adaptive Test (CAT) (Hart et al. 2006) which also adapts to the functional level of the individual patient. A notable strength of the FLEXSF and CAT are their ability to tailor the mechanism for assessing patients according to their different baseline levels, therefore potentially addressing concerns regarding using the same measure with different target populations such as high level sports people versus the frail (Schmidt et al. 2014).

Engebretsen et al. (2010), Mintken et al. (2010), Kromer et al. (2014), and Chester et al. (2016) also used a shoulder specific functional outcome score, namely the Shoulder Pain and Disability Index (SPADI) which comprises 5 pain related questions and 8 disability related questions. The SPADI provides three scores: the total pain score, the total disability score and the total SPADI score of which Kromer et al. (2014) chose only to use the disability sub-score. The high reliability, validity, and responsiveness of the SPADI have been previously documented (Roach et al. 1991; Heald et al. 1997) making this a robust measure of shoulder specific function. Indeed the SPADI has been widely tested for internal reliability (Roddey et al. 2000) and test-retest reliability (Beaton and Richards 1998). Its construct validity in relation to other shoulder function scores has been identified as being high (Angst et al. 2004; Cloke et al. 2005). Furthermore, effect sizes in relation to shoulder pain managed with physiotherapy have been reported (Heald et al. 1997). Alongside the SPADI, Chester et al. (2016) also used the Quick-DASH (Disabilities of the Arm, Shoulder and Hand Score) and Kennedy et al. (2006) used the DASH which are both robust measures of upper limb specific function (Desai et al. 2010) although less specific to the shoulder.

The baseline scores for the above outcome measures were explored as prognostic factors but Chester et al. (2016) and Kennedy et al. (2006) also used the final scores as the dependent variables for determining outcome. A similar approach was taken by Engebretsen et al. (2010), Kromer et al. (2014) and Chester et al. (2016) with the SPADI, Bartolozzi et al. (1994) with the UCLA and Deutscher et al. (2009) with the CAT. It can be considered that where the same tool is used as both a potential predictor and also as a determinant of outcome this is methodologically questionable due to the measurement overlap between one of the independent variables and the dependent variable.

It is therefore concluded that the assessment of function as a potential prognostic factor has merit so that the potentially wide-ranging impact on outcome can be considered. It is advocated that a fit for purpose, shoulder specific assessment of function is used and applied in its entirety rather than only a component thereof. However the large variation in outcome measures used underlines both uncertainty as to which outcome measure is best suited and the wider issue of lack of consensus regarding shoulder pain and specifically SIS/RCTendinopathy. Furthermore, the shoulder specific assessment of function should not be the same measure that is also used to determine patient outcome.

2.5.6 D1. Clinical measures: Strength

The amount by which a patient with SIS/RCTendinopathy is able to generate force using their shoulder is relevant because strength impairment +/- ROM impairment can result in functional limitation (Payne and Michener 2014). Furthermore strengthening the muscles of the shoulder, particularly the rotator cuff, are major components of non-invasive, multimodal physiotherapy (Holmgren et al. 2012).

In the context of prognosis, it could be postulated that substantial impairment of strength at baseline could be associated with worse outcome due to the high pain levels and/or pathology-related deconditioning that might underpin it (Cook and Purdam 2009). Such patients might be theorised to require substantial correction of their strength impairment for them to be classified or self-identify as having improved. Indeed where the rotator cuff is torn, it might be perceived that any associated weakness might not be correctable via non-invasive, multimodal physiotherapy alone.

The study by Bartolozzi et al. (1994) used manual muscle testing via a grading system to quantify their subjects' strength. Despite grading out of five, they chose to categorise the scores according to (i) no weakness, (ii) mild weakness if 4/5, (iii) moderate weakness if 3/5 and (iv) severe weakness if the patient was unable to actively elevate the arm against gravity. However they failed to provide evidence of the reproducibility of the measure and so confidence in the strength-related evidence from this study is therefore low.

In the study by Kennedy et al. (2006) a combination of muscle wasting and decreased muscle strength were recorded as being mild, moderate, or severe. However for analysis purposes the authors chose to collapse down these ratings to 'some degree of decreased strength' versus

'normal'. As with the other grading system, such an approach has the advantage of incorporating dynamic, through-range movements, thereby ensuring clinical and functional relevance. However whilst the authors' rationale was that dichotomising the data might address potential issues around inter-rater reliability, they failed to provide operational definitions for the ratings or assess the reliability of how the ratings were applied. The robustness of the strength-related evidence from this study is therefore limited.

Mintken et al. (2010) reported muscle strength as 'weakness' or 'no weakness', using the definition for specific muscles from the textbook by Kendall et al. (1993). The reporting by Mintken et al. (2010) of moderate to substantial agreement between testers for most of the muscles they tested provides evidence for the reliable application of this tool. However the use of binary measures by Kennedy et al. (2006) ('some degree of decreased strength' versus 'normal') and Mintken et al. (2010) ('weakness' versus 'no weakness') mean that their measurement tools arguably have limited discriminative capacity.

In contrast to the above studies, Hung et al. (2010) and Chester et al. (2016) took quantitative measures of muscle force using hand-held dynamometers. Hung et al. (2010) used a 'make test' whereby an isometric contraction is generated by the subject against a fixed point, namely the assessor holding the dynamometer (Conable and Rosner 2011). However as with the manual muscle testing, the accuracy of this approach is dependent upon the assessor being able to fully resist the subject's force. Nonetheless hand-held dynamometry offers an accurate measure of the force generated with units such as the one used by Hung et al. (2010) having manufacturer reported accuracy levels of 1% of the 0-300 lbs range and the ability to record in 0.1 lbs increments (http://www.hogganhealth.net/pdfs/microfet2.pdf; last accessed 24/04/2017).

However for research applications to be meaningful the tool must be applied in a reproducible manner. Chester et al. (2016) provided no details of any reliability assessment. This is particularly relevant because the clinical measures in the study by Chester et al. (2016) were collected by treating clinicians across 11 NHS trusts and social enterprises. However no details of individual or collective reliability for the measures were presented. Uncertainty as to the inter- and intra-rater reliability of the strength measures therefore raises questions as to the accuracy of the reported strength measures. Hung et al. (2010) cited the intra-rater reliability data from a paper by McClure et al. (2006). However the McClure et al. (2006) intra-rater reliability data actually referred to a different type of muscle test, the 'break test' whereby the subject was resisted until they are able to

'break' the resistance offered by the tester. This raises the spectre of lower confidence in the strength data reported by Hung et al. (2010).

The specific muscles and movements assessed in prognostic studies varied considerably. Where grading systems were used some authors stated the specific muscles and physiological movements (Serratus Anterior, Middle trapezius, Lower trapezius, Rhomboid, Deltoid, External and internal shoulder rotator muscle strength by Mintken et al. 2010), others assessed only specific muscles (supraspinatus, infraspinatus, serratus anterior by Kennedy et al. 2006) and some failed to describe either the specific muscles or physiological movements (Bartolozzi et al. 1994). Deriving a consensus from such disparate approaches is therefore challenging.

Conversely where studies used hand-held dynamometry there was a degree of consistency, with Hung et al. (2010) assessing shoulder abduction, internal rotation (IR), external rotation (ER) force and scapular protraction force and Chester et al. (2016) measuring abduction and ER. The consistent use of ER mirrors the concept of posterior rotator cuff strength (Reinold et al. 2004). However both Hung et al. (2010) and Chester et al. (2016) failed to describe their testing positions which is of relevance because it can be postulated that strength measures will be highly influenced by which part of the muscle range the testing is performed in and also whether the test is performed in a position where pain is present due to potential pain inhibition. In addition the use of composite planes of movement such as scaption are frequently used in the literature due to the functional nature of the plane and the ability for it to partially represent both abduction and flexion, whilst avoiding excessive burden of testing (Hsu et al. 2009).

It can be concluded that a variety of approaches including different measurement tools and aspects of strength including different shoulder muscles and movements have been assessed in previously published prognostic studies. Whilst grading systems offer arguably greater clinical and functional relevance than measures of static muscle strength via dynamometry, isometric force measures offer greater sensitivity to change and reproducibility, thereby providing more robust research data. However the reporting of application-specific reproducibility data should be a pre-requisite along with standardisation of the limb position during testing.

2.5.7 D2. Clinical measures: ROM

The amount by which a patient with SIS/RCTendinopathy is able to move their shoulder is relevant because ROM impairment can result in functional limitation and so restoration of ROM is both a

major component of non-invasive, multimodal physiotherapy and a key variable used to assess outcome (Holmgren et al. 2012; Payne and Michener 2014). In the context of prognosis it could be postulated that substantial impairment of ROM at baseline could be associated with worse outcome due to the high pain levels, deconditioning or structural / functional deficits that might underpin it (Koester et al. 2005)). Such patients might be theorised to require substantial correction of the ROM impairment for them to be classified or self-identify as having improved.

Hung et al. (2010) and Tyler et al. (2010) both measured passive GHJ IR and ER ROM in supine with the arm in 90° abduction using a goniometer and inclinometer; respectively. Measuring passive GHJ ER ROM can be used as an exclusion criterion for patients with adhesive capsulitis (Vermeulen et al. 2006). However Tyler et al. (2010) assessed passive GHJ IR and ER ROM as a potential contributing factor to posterior shoulder tightness which has been theorised to be linked to internal impingement (Spiegl et al. 2014). As part of this Tyler et al. (2010) and Hung et al. (2010) also measured posterior shoulder tightness (PST) and glenohumeral internal rotation deficit (GIRD) again via an inclinometer. Mintken et al. (2010) quantified PST by measuring cross-chest adduction in supine along with passive ROM of the shoulder in respect to shoulder abduction and both IR and ER at 90° abduction. Chester et al. (2016) measured passive flexion, abduction and ER but gave no details as to the testing positions.

Compared to passive ROM, active ROM is arguably more relevant to SIS/RCTendinopathy due to the movement-related nature of the pathology (Koester et al. 2005). It is therefore unsurprising that the majority of studies assessed active ROM, including those that assessed composite or functional movements. Engebretsen et al. (2010) assessed hand behind back ROM whilst Mintken et al. (2010) used a battery of three functional tests. For Mintken et al. (2010) these were hand to neck, hand to scapula and hand to opposite shoulder. Engebretsen et al. (2010) quantified the ROM by recording the maximal position of the patient's thumb to the corresponding vertebral level. Whilst such approaches have the advantage of being quick to perform, potential errors around spinal level palpation threaten the robustness of this measure. In contrast, the approach used by Mintken et al. (2010) involved grading each movement on a scale of 0-4 using descriptors relating to the gross amount and quality of movement (Yang and Lin 2006). Such a composite assessment is highly clinically relevant; however the authors failed to assess the reproducibility of how this measure was applied. This is particularly relevant given that they had nine different assessors and so the accuracy of their data could be questioned.

Confidence in the data reported by Kennedy et al. (2006) was even more questionable as they failed to provide operational definitions for their criteria 'mild, moderate or severe' impairment of ROM. This was particularly relevant given the potential for variation in interpretation between the 81 assessors who participated in the project. Furthermore they failed to state which movements (e.g. planar or functional movements) the ROM data related to, thus severely undermining the value of their data.

Other studies which assessed active ROM considered planar movements such as flexion (Bartolozzi et al. 1994; Engebretsen et al. 2010; Mintken et al. 2010; Chester et al. 2016) and abduction (Bartolozzi et al. 1994; Chester et al. 2016). In terms of symptomatic movements, active flexion and abduction can be considered relevant movements to assess in SIS/RCTendinopathy due to the frequency with which patients report symptoms during shoulder elevation (Koester et al. 2005). However scaption can be considered a preferred movement to asses as it is a composite planar movement, occurring at 30-40° anterior to the frontal plane (Hsu et al. 2009). Thus it has the advantage of assessing GHJ ROM where the scapula lies in its natural position on the thoracic wall, thereby replicating a more typically functional movement plane (McClure et al. 2006).

The accuracy of both the tool and its application by an assessor influence the reliability of the subsequent ROM findings. In measuring bilateral flexion, Engebretsen et al. (2010) used an inclinometer and cited an accuracy of 5°, although no reference for this figure (neither manufacturer's data nor their own reproducibility study) could be located. Mintken et al. (2010) reported shoulder flexion measured in degrees via goniometry but provided no detail of its accuracy. Bartolozzi et al. (1994) failed to state what measurement tool was used. Chester et al. (2016) provided no evidence of inter or intra rater reliability for the clinicians who collected the ROM measures in their multi-centre investigation. This was of particular concern considering the poor inter-observer agreement in measuring ROM in shoulder pain patients reported by de Winter et al. (2004). There are therefore uncertainties about the accuracy of the ROM data from such studies. One mechanism to address this might be the use of Magneto-Inertial Measurement Unit sensors (MIMU) with high levels of precision reported, although substantial operator training is necessary to attain such accuracy (Bouvier et al 2015).

The issue of ROM relative to pain is of clear relevance in SIS/RCTendinopathy because pain is a frequent reason why patients will report functional restriction (Johansson et al. 2002). Engebretsen et al. (2010) stated that they allowed their patients to decide at which point pain restricted further

movement occurring. This is in contrast to Mintken et al. (2010) who reported the measurement of pain-free active shoulder flexion. Both pain-free and limit of ROM can be considered relevant in SIS/RCTendinopathy because improvement in function related pain is a common therapeutic aim, whilst maximal ROM is often of greatest functional relevance (Johansson et al. 2002).

It can therefore be concluded that a broad range of approaches to measuring ROM have been used in previous prognostic studies. However it is proposed that the measurement of active ROM is more relevant in SIS/RCTendinopathy than passive ROM. The reliability of the tool in the context within which it is used should be assessed and the influence of pain upon the movement should be included.

2.5.8 D3. Clinical measures: Scapular movement and control

The way in which the scapula moves in patients with SIS/RCTendinopathy is potentially relevant because the scapula plays a key role in the control of the shoulder (Lopes et al. 2015). This includes the model of SIS/RCTendinopathy where tissue compression under the acromion is theorised to occur (Atalar et al. 2009). In the context of prognosis, substantial impairment of scapular movement at baseline might be associated with worse outcome due to the theorised correction of scapular control required for a patient to improve. The studies by Hung et al. (2010), Mintken et al. (2010) and Chester et al. (2016) considered scapular movement and control variables as potential predictors.

Chester et al. (2016) used a novel approach, which involved noting the symptomatic response to manual facilitation of the scapula during shoulder elevation (Lewis 2009). This technique could be considered to be a symptom modification test but directly associated with the scapula, which can provide valuable information in guiding clinicians as to the likely benefit of targeting the scapula during treatment.

The prognostic studies by Hung et al. (2010) and Mintken et al. (2010) used a variety of tools. Mintken et al. (2010) used two types of static measures, namely the lateral slide test and scapular index. The lateral scapular slide test (Kibler 1998) is used to identify asymmetrical scapular positioning relative to the thoracic spine. In reporting high inter-rater reliability Intra-class correlations (ICCs) (0.67 to 0.71), Mintken et al. (2010) demonstrated the notable strength of these static measures, namely their high reproducibility. However the lateral scapular slide test can be considered of limited clinical and pathological relevance because it is based upon the assumption

that asymmetry is pathological, when clinical experience and published evidence (Koslow et al. 2003) indicate this to commonly be a normal finding.

The scapular index is a unilateral ratio of sternal notch to coracoid process distance and posterolateral angle of scapula to thoracic spine distance (Borstad 2006). This static measure can be considered by its very nature to be of limited relevance to dynamic shoulder pathology. However Borstad (2006) provided evidence of correlation of the scapular index with both scapular internal rotation and pectoralis minor length – both of which are theorised as contributors to shoulder pathology, thereby providing a degree of clinical relevance.

Alongside quantitative measures, Mintken et al. (2010) assessed scapular function qualitatively as per Kibler et al. (2002). This involved categorising a patient's movement pattern into one of four categories: type I = prominence of inferior scapular angle (excessive anterior tilt), type II = prominence of medial scapular border (excessive internal rotation), type III = prominence of superior scapular border (excessive elevation) and type IV = 'normal', symmetrical scapular motion. Assessment of dynamic, through-range scapular function is a notable strength of such an approach. However this classification approach could be considered to be conceptually flawed because it is based upon mutual exclusivity of movement patterns when in fact scapular function is a composite of planar (elevation and protraction) and rotational (upward rotation, anterior tilt and internal rotation) movements (McClure et al. 2006).

One alternative approach is the scapular dyskinesis test (SDT) which is a qualitative assessment of scapular position and movement based upon descriptive criteria and subsequently rating patients according to normal, subtle or obvious dyskinesis (McClure et al. 2009). Notable strengths include non-mutual exclusivity of movement patterns and validation (albeit using binary dyskinesis ratings) against a three-dimensional movement analysis system (Tate et al. 2009). However it is noted that development and testing was undertaken with high-level athletes and so the transferability of this approach to typical NHS patient populations may be questionable.

Hung et al. (2010) used a different approach via a three-dimensional movement analysis system ('FASTRAK') to quantify scapular kinematics (upward rotation, anterior tilt and internal rotation) at 30°, 60°, 90°, and 120° of humeral elevation in the scapular plane. The objective, through-range measurement of all 3 scapular rotations is a notable strength. Furthermore the inclusion of a weighted trial where subjects held a 2Kg load during humeral elevation enhanced the likelihood of

identifying differences by making use of the observation that scapular dysfunction is more prominent when fatigue or loading is present (Tate et al. 2004). However the substantial time and equipment burden associated with using a three-dimensional movement analysis system makes it prohibitive to use such a tool in clinical practice, thereby limiting the clinical transferability of the findings.

It can therefore be concluded that whilst scapular function as a potential prognostic factor has been assessed in a variety of ways, there are both strengths and limitations with each of the tools previously used. It is proposed that where scapular function is included in prognostic research as part of informing clinical care pathways, any tool used should ideally be reliable, dynamic, conceptually acceptable and clinically applicable in nature.

2.5.9 E1. Structural pathology via imaging

Where a patho-anatomical model of SIS/RCTendinopathy is favoured then the presence and nature of structural pathology in the shoulder is considered highly relevant (Tempelaere et al. 2016). In the context of prognosis, substantial structural pathology at baseline could be theorised to be associated with worse outcome due to the severity of the condition and the assumed inability of non-invasive, multimodal physiotherapy to directly address structural pathology, such as a hooked acromion or rotator cuff calcific deposits or tears. Consequently, such patients might be theorised to be much more likely to fail conservative management. The prognostic studies by Bartolozzi et al. (1994), Morrison et al. (1997) and Ogon et al. (2009) included assessment of structural pathology as a potential prognostic factor.

The study by Morrison et al. (1997) examined acromion morphology using plain X-rays and categorised their subjects' acromions according to being flat, curved, or hooked. Whilst an established categorisation approach was used, the observation of a hooked acromion does not indicate structural pathology *per se*; rather it is a theorised contributor to rotator cuff damage in a compression / impingement model of pathology (Balke et al. 2013). Therefore the absolute relevance to a patient of having a particular acromion type is debatable.

Bartolozzi et al. (1994) used magnetic resonance imaging (MRI) to identify rotator cuff pathology, using the categories of (i) impingement, tendinitis; (ii) partial or a small full thickness tear($<1cm^2$) and (iii) tears $> 1cm^2$. It is unclear how impingement was defined as it is an inherently dynamic concept yet MRI is taken in static postures. Furthermore, 'tendinitis' infers inflammation of the

tendon but it is widely accepted that tendon inflammation is rare and that tendinosis is more commonly observed (Joseph and Denegar 2015). Bartolozzi et al. (1994) provided no further details or reference for their categorising system and so it is not possible to answer these questions.

Ogon et al. (2009) used plain X-rays to identify the presence and nature of any calcific deposits. X-ray has the advantage of being quick and inexpensive. The authors employed two different classification systems (Mole et al. 1993; Gartner and Heyer 1995) for the observed deposits as well as their own system for localising the position of the calcium. They then went on to use diagnostic ultrasound as an additional modality to detail the sonographic appearance of the deposits. Whilst the bi-modality approach used by Ogon et al. (2009) for assessing calcific deposits was thorough, they chose not to report concurrent structural pathology such as rotator cuff tears and subacromial bursitis, thereby prohibiting the inclusion of these potentially important structural pathology findings (Micheroli et al. 2015).

Diagnostic ultrasound can be considered an ideal clinical modality for identifying or excluding rotator cuff tears; whilst for rotator cuff tendinosis and subacromial bursitis it can provide tissue-specific imaging information (Jacobson 2011). When considering SIS/RCTendinopathy then from a patho-anatomical perspective such data could be considered highly relevant (Micheroli et al. 2015). The advantages of ultrasound include the ability to deliver it in the clinical environment, the ability for the sonographer to interact with the patient and for movement of the limb to be incorporated into the assessment (Corazza et al. 2015). However whilst ultrasound is recognised as having comparable sensitivity and specificity as MRI, the highly operator dependent nature of the modality means that extensive training is essential (Ottenheijm et al. 2010). Furthermore, the reproducibility of the operator should be quantified and the uncertain link between structural pathology and symptoms must be acknowledged (Girish et al. 2011).

It can be concluded that a variety of assessment approaches on a range of target tissues and pathologies have been used in previously published prognostic studies. It is proposed that when considering structural pathology, the pathological tissue is imaged rather than theorised through associated structures, such as acromion morphology. Furthermore the imaging of multiple tissues that are potentially involved in SIS/RCTendinopathy is advocated rather than a single element, e.g. calcium deposits. Finally the subjective nature of imaging interpretation means that reproducibility data should be provided for the combination of operator and imaging tool.

2.5.10 E2. Structural pathology via orthopaedic tests

Where a patho-anatomical model of SIS/RCTendinopathy is favoured then, as with diagnostic ultrasound, the presence and nature of structural pathology in the shoulder is considered highly relevant (Calis et al. 2000). In relation to outcome for patients managed with non-invasive, multimodal physiotherapy, the presence of symptomatic "structural pathology" (such as a positive Neer's sign or Hawkins-Kennedy test) could conceptually be used as an identifier of the nature of the pathology (Kromer et al. 2014) and/or used as an objective marker to indicate severity of the condition. As such, it could be theorised to influence and therefore predict outcome.

The studies by Mintken et al. (2010) and Chester et al. (2016) were the only ones to include orthopaedic tests as potential prognostic factors whilst many of the other studies under consideration used them solely as an inclusion criterion.

Mintken et al. (2010) used 13 tests to explore GHJ instability, symptom reproduction from the ACJ and subacromial region along with tests of rotator cuff integrity. This wide-ranging approach reflects the generalised shoulder pain population that the authors investigated. Chester et al. (2016) also investigated a non-specific shoulder pain population but conversely only used a test of rotator cuff integrity, namely the external lag sign. However the combination of tests designed for testing symptom reproduction and structural integrity used by Mintken et al. (2010) did overlap with that of Chester et al. (2016) as Chester et al. (2016) also included the testing of symptomatic response to manual facilitation of the scapula during shoulder elevation. As this later technique is scapular orientated, it was considered earlier in this literature review.

The use of a battery of tests by Mintken et al. (2010) reflects the opinion that singular tests in isolation are of limited value due to issues of poor specificity or sensitivity (Calis et al. 2000). Conversely, in looking to explore structural integrity, the robustness of the single test approach by Chester et al. (2016) would be enhanced by the use of diagnostic imaging.

It can therefore be seen that orthopaedic tests have only been used in a limited capacity in previous prognostic studies considered in this literature review. Where they are applied for the purposes of identifying the symptom reproducing structures then a battery of tests should be used. Recommendations here include standardised performance and interpretation of the findings (Hanchard et al. 2013). A combination of tests with high sensitivity (e.g. Neer's sign, Empty can and Hawkins-Kennedy test) (Alqunaee et al. 2012) and high specificity (e.g. painful arc) (Calis et al. 2000)

are advocated. Furthermore, signs that are commonly used to identify SIS/RCTendinopathy patients such as pain on active abduction (Diercks et al. 2014) have a potential role. A frequently advocated approach is three or more positive tests as a cut-off point, albeit typically for identifying pathology (Michener et al. 2009) rather than prognosis.

It can be concluded that orthopaedic tests are typically used in an attempt to identify pathology and therefore define a sample. However, due to widely identified issues with the sensitivity, specificity and conceptual value of these tests, using them instead as a potential prognostic factor has merit. Of relevance here is the use of a battery of tests and a threshold of three or more positive tests as a cut-off point is advocated.

2.5.11 Conclusion

It can therefore be concluded that a large number of categories of variables have been explored in previous studies and that each of these categories has included often very different mechanisms to quantify the variable of interest. This belies a persistent uncertainty and lack of consensus in this subject area as to the relevant signs and symptoms of SIS/RCTendinopathy in terms of the aetiology, treatment and prognosis of this patient group.

2.6 Treatment / intervention

The wide range of care pathway variants and the treatment approaches available via other professions was briefly described earlier. As noted previously, physiotherapy for SIS/RCTendinopathy includes education and advice, exercise based therapy and hands-on techniques such as manual therapy, taping and electrotherapy. The rationale behind education and advice includes empowering patients, addressing unhelpful belief mechanisms and avoidance of unnecessary symptom exacerbation (Dragesund and Kvale 2016). Exercise therapy (Holmgren et al. 2012; Lewis 2012; Littlewood et al. 2015) may include strengthening of weak muscles due to disuse or pain, including those muscles that stabilise the GHJ, scapula or trunk; also the regaining of ROM and stretching of muscles theorised to place the shoulder into symptom producing positions. Taping (Kalter et al. 2011) is theorised to influence similar mechanisms as well as muscle recruitment (Smith et al. 2009), whilst manual mobilising of the GHJ, cervical or thoracic spine region (Senbursa et al. 2007; Cook et al. 2014) is theorised to facilitate optimal joint glide and local pain inhibition. Electrotherapeutic modalities may be used for pain relief or the theorised addressing of structural pathology via shockwave therapy (Schmitz et al. 2015) although evidence for this modality is lacking.

Additional interventions (Littlewood et al. 2016) may include use of heat or cold, acupuncture and hydrotherapy; whereby pain and muscle spasm relief, ROM and strengthening may be enhanced. To maximise treatment carry over and patient independence and empowerment, home exercise programmes may also be prescribed which may mirror or complement approaches taken in the clinic-based interactions (Bennell et al. 2010). The nature of the treatment approaches used in previous prognostic studies will now be considered.

2.6.1 Types of treatment

The studies by Kennedy et al. (2006) and Sindhu et al. (2012) provided no details of the component elements of the intervention received by the patients in their studies. Ogon et al. (2009) stated they applied a physiotherapy treatment algorithm including heat or cold and manual therapy but provided no further details. It is therefore not possible to determine how representative of physiotherapy the treatments were and so transferability of the prognostic findings is limited.

The studies by Conroy and Hayes (1998), Engebretsen et al. (2010) and Kromer et al. (2014) all used standardised interventions due to their prognostic findings being derived from secondary analyses of RCT data. The use of standardised interventions has the methodological and statistical benefit of controlling for the potential confounding factor of treatment variability thereby providing evidence regarding people who respond differently to specific treatments, i.e. treatment effect modifiers (Kent et al. 2010). However this standardisation places inherent limitations upon the transferability to routine practice of the subsequent prognostic findings.

Subjects in the study by Engebretsen et al. (2010) were treated with either supervised exercises or radial extracorporeal shockwave therapy, where the electrotherapeutic modality used shockwaves to theoretically stimulate pain relief and optimise tissue healing (Harniman et al. 2004). However, such a modality requires specialist equipment which is not routinely used, so the transferability of prognosis when treated with this modality is inherently limited. The other patients in the study by Engebretsen et al. (2010) received supervised exercises which emphasised relearning of normal movement patterns with an initial focus on unloading the rotator cuff via postural correction and manual techniques. Combined with the progression to endurance and control exercises, these are more typical of standard non-invasive, multimodal physiotherapy.

Exercise therapy was a common theme amongst the RCT derived data, with Conroy and Hayes (1998) incorporating muscle strengthening, whilst Kromer et al. (2014) included shoulder girdle and

thoracic spine stretching and strengthening exercises. Manual therapy techniques were also included by these RCTs, with Conroy and Hayes (1998) incorporating joint mobilisation, and Kromer et al. (2014) using manual mobilisation techniques for the shoulder complex and cervical spine. Exercise therapy and manual techniques are considered a main stay of physiotherapy (Bennell et al. 2010) and so their frequent use in these RCTs ensures their findings are broadly representative. However patient education (Dragesund and Kvale 2016), also a mainstay of physiotherapy, was only included by Engebretsen et al. (2010) and Kromer et al. (2014) and only in their intervention arms. The comparability of prognostic findings across each of these studies is therefore potentially limited by the absence for some patients of this critical component of non-invasive, multimodal physiotherapy.

Some studies were not RCTs in nature but nonetheless chose to standardise the interventions received. Mintken et al. (2010) explored prognosis following treatment with only thrust manipulations of the cervicothoracic spine and specifically detailed the prohibiting of clinicians adjusting the intervention, regardless of their clinical findings. A broader approach was taken by Hung et al. (2010) whose standardised regimen comprised stretching, strengthening and ROM exercises along with manual therapy. This ensured a more representative delivery of non-invasive, multimodal physiotherapy, although their standardising of the exercises and even stretch hold times potentially limits the transferability to routine practice.

As with Mintken et al. (2010), Tyler et al. (2010) limited their intervention to components that mirrored their conceptual framework, namely GIRD and PST. As such, they emphasised glenohumeral joint glides, sleeper stretches and cross-chest adduction. Such approaches strengthen the patho-anatomic specificity of the prognostic results and limits the inherent variability introduced by using various treatment approaches. However again this approach limits the transferability to the wider treatment of SIS/RCTendinopathy patients.

A pseudo-controlled approach (where a structured approach was utilised but included elements of adaptability, for example timescales of how long certain treatment components were continued) was taken by Morrison et al. (1997) who used a physiotherapy programme consisting of stretching and then strengthening, whilst Bartolozzi et al. (1994) simply stated they included rotator cuff strengthening and ROM exercises. Virta et al. (2009) similarly detailed the component treatments including ROM and strengthening exercises along with how they were delivered. However in basing the delivery of the regimen upon assessment of the individual patient's dysfunction, they mirrored

routine clinical practice. Kromer et al. (2014) used the novel approach of including an arm in their RCT where alongside a standardised exercise protocol they also incorporated individualised physiotherapy based on clinical examination results. This reflects the complex nature of physiotherapy intervention, tailored to the individual patient. However capturing the specific nature of the treatment can be challenging, as can defining the nature and dose of the intervention. As such, deriving evidence as to the key ingredients of the intervention and its efficacy in different pathologies, settings and patient groups is challenging.

An entirely clinician directed approach was used by Deutscher et al. (2009) and Chester et al. (2016). Such an approach is truly representative of clinical practice where physiotherapists assess and treat their patients using an individualised approach (Kromer et al. 2013). However the challenge posed by this pragmatic approach is the introduction of between-subject treatment variability.

The large-scale study conducted by Deutscher et al. (2009) categorised treatment using an electronic database system (Deutscher et al. 2008). However the clinical utility of this is limited by the requirement to purchase and integrate this electronic system. Chester et al. (2016) on the other hand did not publish details of how the delivered intervention was recorded nor of treatment duration or frequency and so it is not possible to identify the representativeness or otherwise of the treatment received by patients in their study.

2.6.2 Duration and frequency of treatment

Wide variation in the duration and frequency of treatment was seen in the previously published prognostic studies considered in this review. Where studies used a standardised approach and provided details of treatment duration this ranged from a matter of days (Mintken et al. 2010) to 3 weeks (Conroy and Hayes 1998), 5 weeks of clinician directed non-invasive, multimodal physiotherapy followed by 7 weeks of home based exercises (Kromer et al. 2014), 6 weeks (Hung et al. 2010), 8 weeks (Virta et al. 2009), a maximum of 12 weeks (Engebretsen et al. 2010) and a minimum of 3 months (Ogon et al. 2009). In terms of frequency of treatment, those who reported data ranged from once a week (Engebretsen et al. 2010), twice a week (Virta et al. 2009; Hung et al. 2010; Kromer et al. 2014) to 3 times per week (Conroy and Hayes 1998; Tyler et al. 2010).

The inherent variability in the subsequent intensity and dose of non-invasive, multimodal physiotherapy received in these different studies is a limiting factor in pooling prognostic findings across them. Although Kennedy et al. (2006) provided no details of content, they reported a mean of

15 sessions (SD=9) over a mean of 65 days (SD=26), whilst Deutscher et al. (2009) reported a mean of 9 sessions (SD=6) over a mean of 56 days (SD=40). The broadly comparable number of sessions and duration in these studies could be viewed as a providing a benchmark for studies such as Chester et al. (2016) who also used a pragmatic approach. However location of the clinical setting (Israel, Canada and UK for Deutscher et al. (2009), Kennedy et al. (2006) and Chester et al. 2016; respectively) along with local funding pressures (private versus public health) may also influence the availability of healthcare resources including the number and timing of physiotherapy appointments.

In conclusion, to optimise the reproducibility and accurate interpretation of prognostic study findings, details of the intervention delivered, including duration and frequency should be recorded. Clinician led treatment is arguably the most representative but brings challenges in terms of capturing the nature of the intervention delivered and it potentially introduces additional variability into a study.

2.7 Response to intervention

The proportion of subjects in previous prognostic studies who did and did not get better is of relevance both from the perspective of prognostic statistical analysis and also as a reflection of the uncertainty as to which patients do and do not respond to particular treatment approaches. The non-RCT studies which reported rates of successful outcome ranged from 61% (Mintken et al. 2010), 67% (Morrison et al. 1997), 70% (Hung et al. 2010), 73% (Ogon et al. 2009) to 87% (Virta et al. 2009). Ironically, the highest successful outcome rate (Virta et al. 2009) was for those patients on a surgical waiting list.

In terms of patients worsening with physiotherapy, these rates were typically low. Where reported, these ranged from 3% (Engebretsen et al. 2010), 4% (Mintken et al. 2010) to 7% (Kennedy et al. 2006).

2.8 How outcome is defined and measured

The mechanism by which outcome is measured and defined are key considerations in the design of a prognostic study. The outcome might be mortality or survival across a specific time period; escalation of treatment or failure with a particular treatment approach; through to levels of symptoms and function or returning to work (Altman et al. 2009). The choice of outcome will therefore have a substantial impact on the subsequent findings and how they are interpreted. Similarly, the time scale over which outcome is measured and whether the final state or the change

in state from baseline is the dependent variable, all will have similar importance. In the area of shoulder prognosis, areas of variation include mechanisms such as clinical judgements of symptom change or treatment escalation, measurement of an outcome as a continuous variable or collapsing down to a categorical or binary variable.

In considering prognostic studies relevant to this thesis, authors such as Ogon et al. (2009) used treatment-related outcomes. Ogon et al. (2009) defined a negative outcome by progression to advanced therapeutic measures, which included extracorporeal shock wave therapy or surgery. A strength of this approach is that it reflects an escalation along the clinical pathway, including a likely increased cost to the healthcare provider (Carr et al. 2015). However a limitation is that a decision to escalate the intervention may be influenced by clinician preference and so may be subject to bias or it may not be aligned to the patient's perception of outcome. Furthermore as a binary decision, it arguably lacks sensitivity. Alternatively Engebretsen et al. (2010) included working status as an outcome measure. This has limitations regarding those who are studying, unemployed or retired. However this choice of outcome measure means that the study findings can provide valuable information for the patient (in terms of financial and job security), their employers and wider society when considering taxation and productivity, sickness or unemployment benefit costs (Linaker and Walker-Bone 2015).

The other studies under consideration all used patient based measures. The study by Conroy and Hayes (1998) simply used a VAS pain score as its determinant of outcome. However the breadth of the patient experience and impact of shoulder pain is typically multifactorial where pain is only one component (Gartsman et al. 1998). In order to address this, the majority of prognostic studies in this area have used validated, patient-reported outcome measures to determine outcome. Hung et al. (2010) and Mintken et al. (2010) both used the GROC (Jaeschke et al. 1989) which is a 15-point global rating scale ranging from -7 ('a very great deal worse') to 0 ('about the same') to +7 ('a very great deal better'). Both authors chose to use a score of +4 or greater (i.e. 'moderately better' and above) to define in a binary form a positive outcome. However such an approach is inherently subjective to recall bias, as it requires patients to rate how much they perceive they have changed relative to baseline.

An alternative approach is the use of outcome measures completed both at baseline and at the follow up time point(s). Kennedy et al. (2006) chose to use the DASH to measure disability whilst Chester et al. (2016) used the QuickDASH. There is evidence in the literature (Turchin et al. 1998;

Beaton et al. 2001) that this measure is reliable and valid. However a shoulder-specific outcome was used by the majority of other authors. Bartolozzi et al. (1994), Morrison et al. (1997) and Virta et al. (2009) all used the UCLA, which covers pain, function and overall satisfaction alongside physical measures of active ROM and strength. In contrast Deutscher et al. (2009) and Sindhu et al. (2012) used purely patient reported outcomes via the shoulder CAT (Hart et al. 2006). This robust outcome measure (Crane et al. 2006; Hart et al. 2010; Wang et al. 2010) has the advantage of adapting to the functional level of the individual patient in assessing their functional status in terms of ability to perform daily tasks.

Tyler et al. (2010) used the Simple Shoulder Test (SST) which comprises 12 questions regarding shoulder function. Whilst robustness of the tool has been reported in the literature (Beaton and Richards 1998; Roddey et al. 2000; Godfrey et al. 2007) the timescale over which the questions relate to (i.e. at the moment of assessment) limits the ability to representatively capture the sometimes fluctuating nature of symptoms.

The SPADI was used by Engebretsen et al. (2010), Kromer et al. (2014) and Chester et al. (2016) to determine outcome. As noted in the section on function / disability, the SPADI can be considered a robust measure of shoulder specific function. However the methodological concerns regarding the use of the same tool as a potential prognostic factor and as a determinant of outcome have been discussed earlier. One mechanism to partially overcome this is to use a different functional outcome score for the dependent variable compared to that explored as a potential prognostic factor. One such tool is the Oxford Shoulder Score (OSS) (Dawson et al. 1996) which is highly rated for its psychometric properties and has the advantage of being developed on a cohort of shoulder patients largely comprising those with impingement symptoms (Desai et al. 2010; Ekeberg et al. 2008). There is published evidence of the internal consistency and test-retest reliability of the OSS (Huber et al. 2004). Minimal clinically important change data specific to patients with rotator cuff disease has been published (Ekeberg et al. 2010b). As the OSS has been widely used in the surgical literature then given the potential for patients with SIS/RCTendinopathy who receive physiotherapy to potentially escalate their treatment to surgery (Butt et al. 2015), the use of such an outcome could enhance the transferability of the findings between different professional and clinical elements of the care pathway.

Where authors have used outcome scores, their handling of the variable is highly relevant. In this respect Bartolozzi et al. (1994), Morrison et al. (1997), Kennedy et al. (2006), Deutscher et al. (2009),

Engebretsen et al. (2010) and Chester et al. (2016) used the final scores as their dependent variable irrespective of baseline. Tyler et al. (2010) took this one step further and dichotomised their patients (based upon their SST scores) into those with and those without residual symptoms. However the use of only final scores fails to reflect any actual change that has occurred over the treatment and follow up period and so can be considered of limited relevance beyond reflecting final state. Conversely Kennedy et al. (2006), Sindhu et al. (2012) and Kromer et al. (2014) calculated a change score (including direction) from baseline, which has the advantage of incorporating baseline state.

2.8.1 Timing

The time scale over which outcome is measured plays a role in determining the applicability of the subsequent prognostic information. The studies by Conroy and Hayes (1998), Deutscher et al. (2009), Virta et al. (2009), Hung et al. (2010), Mintken et al. (2010), Tyler et al. (2010) and Sindhu et al. (2012) all collected their outcome scores upon completion of treatment. This provides meaningful information with regards to whether a patient will respond by the end of treatment. However if only this time point is used it provides very limited information regarding the stability of any change over time. The study by Kennedy et al. (2006) used a broadly comparable approach except that discharge data collection occurred whichever happened first: (i) when the patient was discharged or (ii) after a 12-week period of treatment, thereby introducing inherent variability regarding the timescale of data collection relative to the completion of face-to-face treatment.

The studies by Morrison et al. (1997) and Bartolozzi et al. (1994) used longer but again non-standardised follow up time points, with means of 27 months (range 6 to 81 months) and 20 months (range 6 to 41 months); respectively. Whilst such timescales provide valuable information regarding stability of change over time, the range of follow up time points introduces inherent variability as a potential confounding factor. Conversely, the study by Engebretsen et al. (2010) used a standardised time point of 1 year. However as treatment in this RCT was 4-6 weeks and up to 12 weeks depending on the treatment arm then there was inherent variability within the cohort in terms of the time between discharge and follow up. Kromer et al. (2014) used a 3 month follow up time point and Chester et al. (2016) used a combination of short (6 weeks) and longer (6 months) term follow up. However, their use of these time points from the *start* of physiotherapy means that conceivably a large number of patients at particularly the 6-week and 3-months time points were still under treatment, thereby introducing this as a potential confounding factor. Nonetheless the use of a 3-month post-discharge time point has the potential to provide meaningful information regarding maintenance of any change over time. However it is proposed that anchoring the follow up periods

to the completion of face-to-face treatment or discharge from physiotherapy will ensure maximum clinical relevance of the subsequent findings.

2.8.2 Loss to follow up

The studies by Bartolozzi et al. (1994), Morrison et al. (1997), Kennedy et al. (2006), Ogon et al. (2009), Virta et al. (2009) and Hung et al. (2010) stated their loss to follow up rates which varied from 3% (Hung et al. 2010) to 26% (Virta et al. 2009). However the above studies failed to compare the baseline characteristics of those who were and those who were not followed up. Analysis of baseline differences between these two groups is essential in order to identify any bias within the final analysis group. Consequently the potential for such bias in the results from these studies cannot be discounted.

Deutscher et al (2009), Engebretsen et al (2010), Sindhu et al (2012) and Chester et al (2016) reported loss to follow up rates between 10% (Engebretsen et al. 2010) through to 43% (Sindhu et al. 2012) and 60% (Deutscher et al 2009). All of these authors undertook an analysis of any differences between groups, with Engebretsen et al. (2010) reporting their missing subjects as being older and with a higher level of functional impairment at baseline compared to those for whom follow up at 1 year was possible. Sindhu et al. (2012) also identified differences between those available for and lost to follow up in terms of age, pain levels and function but the authors failed to provide details on the direction of difference. Conversely Chester et al. (2016) reported younger patients as being lost to follow up as well as those not partaking in leisure time physical activity. Deutscher et al. (2009) reported their missing cohort subjects as comprising a higher percentage of patients with a chronic condition at intake and a higher number of comorbidities. As each of the variables reported by these authors might conceptually impact on prognosis then the reporting of such baseline differences is essential so that any potential bias can be acknowledged.

2.8.3 Blinding and potential bias

Few studies reported whether sources of potential bias were addressed and the nature of how blinding was achieved. Where the researcher was also the person delivering treatment (as in Conroy and Hayes 1998) then the potential for conscious or unconscious bias must be acknowledged. Conversely Hung et al. (2010) explicitly stated that the assessor was blinded to treatment.

The approach used by Chester et al. (2016) had the benefit of the investigator not being involved with the face-to-face collection of any of the variables and so intentional or unintentional bias was

removed. However the dependent variable in the study by Chester et al. (2016) was the absolute outcome scores at 6-weeks and 6-months. As such, blinding of the assessor to the dependent variable may not have been possible.

In conclusion, there was wide variation in the type of outcome measures used and the timescales over which they have been applied in the shoulder prognostic studies considered in this chapter. However the use of a shoulder specific, functional outcome score has the advantage of being condition specific and providing robust, meaningful information to inform prognosis. The use of a change score as the outcome measure has the advantage of reflecting change from baseline state. The timescale over which it is applied should be anchored to the actual end of treatment, but also be long enough to capture any sustained change beyond cessation of treatment.

Loss to follow up should be reported along with statistical analysis of any differences between those who were followed up and those who were not. Furthermore the assessor should not be involved with the treatment of the patients and should be blinded to the individual patient's outcome until all other data collection and processing is completed.

2.9 Analysis approach

As noted in the introduction chapter, prognostic research involves the analysis of multiple and potentially disparate factors to explore their individual and collective influence on outcome (Royston et al. 2009). This approach aligns well with the likely multifactorial nature of SIS/RCTendinopathy (Hanchard et al. 2013).

2.9.1 How candidate variables are selected, including the number of cases per variable

As demonstrated earlier in this chapter, a large number of candidate variables across a broad range of categories have been used in previous prognostic studies. However a critical element is the threat of over-fitting in prognostic models where too many candidate variables are explored in a given sample size (Moons et al. 2009b).

The studies by Bartolozzi et al. (1994), Morrison et al. (1997), Conroy and Hayes (1998), Ogon et al. (2009), Virta et al. (2009), Tyler et al. (2010) and Chester et al. (2016) made no attempt to trim the number of candidate variables. However other studies reduced the number of candidate variables via statistical methods by looking at univariate relationships between variables. The studies by Kennedy et al. (2006), Engebretsen et al. (2010), Hung et al. (2010) and Mintken et al. (2010) tested

any relationship between potential prognostic factors and the dependent variable using a liberal significance level (p<.10) so as to avoid excluding potential predictive variables (Cohen et al. 2013). However such an approach is widely cautioned against due to the failure to consider confounder interaction (Sun et al. 1996), particularly with small data sets (Steyerberg et al. 2001).

Collapsing of variables within a category was performed by Kennedy et al. (2006) if distributions were too low in any one variable. A similar approach was taken by Deutscher et al. (2009) who chose to exclude variables where the incidence was too low. Collinearity amongst independent variables was assessed by Kennedy et al. (2006), Deutscher et al. (2009), Engebretsen et al. (2010) and Kromer et al. (2014). Where sufficient correlation was identified in these studies, the variable with the greatest clinical relevance was retained.

The combining of a statistical and conceptual approach to variable selection was particularly evident in the study by Kromer et al. (2014) who applied *a priori* a targeted approach to the selection of potential prognostic factors relative to their final sample size. Specifically they collected a full range of demographic and clinical-related variables but only entered into their prognostic model those that they were interested in from a theoretical perspective. This allowed them to comprehensively define their sample but select only an appropriate number of key variables relative to their sample size for the subsequent prognostic statistical analysis.

The resulting number of cases per variable in the studies was highly variable. These ranged from 0.5 (Hung et al. 2010) to 5.5 (Engebretsen et al. 2010), 6 (Mintken et al. 2010), 7 (Tyler et al. 2010), 10 (Bartolozzi et al. 1994; Kennedy et al. 2006), 11.8 (Chester et al. 2016), 18 (Virta et al. 2009), 25 (Ogon et al. 2009) and 123 (Morrison et al. 1997). Where a minimum of 10 cases per prognostic factor is advocated (Peduzzi et al. 1996) then the findings from several of these studies are unlikely to be replicated in a new sample due to the high risk of overfitting.

2.9.2 Multivariate regression analysis approach, including logistic versus linear regression

Multivariate regression techniques are typically employed in prognostic research to examine the interaction of multiple variables in predicting a particular outcome (Hemingway et al. 2009). However there is a lack of consensus as to the most appropriate approach for prognostic model building (Royston et al. 2009).

One method is a hierarchical regression model as used by Kromer et al. (2014) whereby following their first step (which concerned variables they were controlling for) their second step explored clinical variables and their third step explored psychological variables. A benefit of this approach is the ability to investigate the contribution of multiple variables within a category before moving onto the next category. However the selection and ordering of the specific categories introduces potential bias into the subsequent model (Field 2009).

Another source of potential bias is demonstrated by the decision by both Engebretsen et al. (2010) and Kromer et al. (2014) to retain numerous variables in their model, irrespective of their contribution or level of significance, thereby adjusting their model for these factors. For both studies, age and gender were controlled for whilst Engebretsen et al. (2010) also included treatment group. As the Engebretsen et al. (2010) study was examining data from a two arm RCT then controlling for treatment group has merit. However the use of other variables has the potential to undermine confidence in the final prognostic model due to the arbitrary retention of candidate variables whose interaction with other factors may or may not be of relevance. A stepwise but non-hierarchical strategy was employed by the majority of other studies, whereby all permutations of variables can be considered, irrespective of the category to which they belong (Royston et al. 2009).

In terms of constructing the most parsimonious model, studies such as Kennedy et al. (2006), Ogon et al. (2009), Engebretsen et al. (2010) and Kromer et al. (2014) used a simple approach whereby variables with $p \geq 0.05$ were removed. However a combined forward selection and backward elimination approach as used by Deutscher et al. (2009), Hung et al. (2010) and Chester et al. (2016) has the advantage of ensuring that statistically justified removal and retention of variables are performed so as to derive the most accurate model (Royston et al. 2009). An example of the approach used comes from Hung et al. (2010) where a significance of $p \leq 0.05$ was required to enter a variable into the model and a significance of $p \geq 0.10$ was required to remove it.

The use by studies of either logistic or linear regression was determined by the nature of their dependent variables as binary or continuous; respectively. As noted previously, the nature of the outcome variable has substantial implications for the interpretation of the subsequent prognostic findings.

Where studies have used outcomes which are inherently binary in nature such as yes/no progression to advanced therapeutic measures by Ogon et al. (2009) or yes/no in work by Engebretsen et al.

(2010) then logistic regression was undertaken. Other studies chose to convert actual or potentially continuous variables into binary outcomes, such as Hung et al. (2010) and Mintken et al. (2010) who both used the GROC to define in a binary form a positive outcome. However such an approach inevitably leads to a loss of statistical sensitivity (Altman and Royston 2006) and so where possible the use of a continuous variable is preferred.

The subsequent reporting of R^2 values indicate the amount of variation in the outcome variable explained by each model. Depending on the reference outcome these ranged from 23% (Kennedy et al. 2006) to 30% (Deutscher et al. 2009; Engebretsen et al. 2010; Chester et al. 2016), 34% (Chester et al. 2016), 36% (Deutscher et al. 2009), 43% (Chester et al. 2016) and 48% (Kromer et al. 2014; Chester et al. 2016).

However some studies such as Bartolozzi et al. (1994), Ogon et al. (2009), Virta et al. (2009) and Hung et al. (2010) failed to present their R^2 values or regression equation coefficients (β). β are used to identify the direction of any relationships between predictor variables and the dependent variable and the significance of the β values are used to identify the magnitude of the contribution of the variable. As such, consideration of the contribution from each variable to the overall model was not possible to establish from these studies nor the overall percentage of variability explained by the combination of factors. These highlight serious flaws in the standard of reporting in prognostic research (Hemingway et al. 2009).

In conclusion therefore, when examining predictors of outcome in SIS/RCTendinopathy a multivariate analysis approach is advocated where the number of candidate variables is aligned to the sample size in order to address the threat of over-fitting. An *a priori* approach to variable selection should be used incorporating one or more of: exclusion of low incidence variables, collapsing of variables within a category and assessment of correlations to select the variable with the greatest clinical or conceptual relevance. Where possible, the outcome variable should be retained as a continuous variable and a stepwise multiple regression approach using forward selection and backward elimination should be used. Subsequent β and R^2 values should be reported.

2.10 Overall quality of the studies

There is merit in attempting to identify a hierarchy of studies from those considered in this literature review. However there is a limited consensus on mechanisms by which to assess the quality of prognostic study evidence (Altman 2001; Hayden et al. 2006). Previous systematic reviews (Kuijpers

et al. 2004; Hayden et al. 2006; Chester et al. 2013) have used bespoke checklists. Alongside the formal assessment of risk of bias and applicability via the PROBAST process, an overall judgment on the type of study design was undertaken to complement the project specific critique already undertaken. In this respect the hierarchy of: (i) prospective cohort study, (ii) RCT derived data, (iii) retrospective cohort study, was applied (NHMRC 2000) where Phillips et al. (2016) specify >80% follow up as an additional criterion for a prospective cohort study.

Applying the above criteria, the studies by Ogon et al. (2009), Hung et al. (2010), Mintken et al. (2010), Tyler et al. (2010) and Chester et al. (2016) (for their 6 week follow up dataset) were identified as being of the highest level of quality as they were prospective cohort studies with >80% follow up. Next were the studies by Deutscher et al. (2009), Kennedy et al. (2006), Virta et al. (2009) and Chester et al. (2016) (for their 6 month follow up dataset) as they were prospective cohort studies but with <80% follow up. The RCT derived data studies by Conroy and Hayes (1998), Engebretsen et al. (2010) and Kromer et al. (2014) were the next lowest. Finally the lowest level of quality were the retrospective cohort studies by Bartolozzi et al. (1994), Morrison et al. (1997) and Sindhu et al. (2012).

2.11 Prognostic findings from previously published papers

The factors identified from previously published papers as being prognostic or not, and the direction of prediction, will now be presented. As with section 1.5 where potential prognostic factors and how they were measured was considered, the gross categories of (A) demographics, (B) clinical history, (C) patient reported measures, (D) clinical measures and (E) structural pathology will be used.

2.11.1 A. Demographics

Table 2-2: A. Demographics factors, segregated according to (i) Generic demographic variables and (ii) Work, recreational or functional task related variables

Paper	Prognostic factor; *direction*	Not prognostic
	(i) Generic demographic variables (ii) Work, recreational or functional task related variables	**(i) Generic demographic variables (ii) Work, recreational or functional task related variables**
Bartolozzi et al. (1994)		**(i)** • Age • Gender

		(ii) • Occupation
Chester et al. (2016)	**(i)** • Gender: being female rather than male; *worse disability via absolute Quick DASH score at 6 weeks post baseline* **(ii)** • Work status: employment: not being in employment due to redundancy, unemployment or disability, compared with being in employment or education; *worse pain and disability via absolute SPADI score at 6 weeks* and *both worse pain and disability via absolute SPADI score and worse disability via absolute Quick DASH score at 6 months post baseline* • Physical activity levels: most strenuous weekly exercise classified as 'none' compared to 'moderate'; *worse pain and disability via absolute SPADI score at 6 months post baseline*	**(i)** • Age • Social deprivation: index of multiple deprivation • BMI • Smoker status **(ii)** • Work status: currently off work due to shoulder pain • Occupation: nature of employment • Occupation: type of work or regular activity
Conroy and Hayes (1998)		**(i)** • Age
Deutscher et al. (2009)	**(i)** • Gender: being female; *poorer functional status via absolute CAT score at discharge* • Age: older age; *poorer functional status via absolute CAT score at discharge*	**(i)** • Smoker status **(ii)** • Physical activity levels
Engebretsen et al. (2010)	**(i)** • Educational attainment: low education; *worse pain and disability via absolute SPADI score and 'not working' status at 12 months post baseline*	**(i)** • Age • Gender **(ii)** • Work status • Occupation: frequency of working above shoulder height • Occupation: frequency of carrying heavy loads at work

Hung et al. (2010)		**(i)** • Age
Kennedy et al. (2006)	**(i)** • Gender: female; *greater disability via absolute DASH score at discharge / 12 weeks* • Age: older age; *greater disability via absolute DASH score at discharge / 12 weeks* • Age: younger age; *greater improvement in disability via change from DASH score at baseline to discharge / 12 weeks* **(ii)** • Work status: workers compensation claim; *greater disability via absolute DASH score at discharge / 12 weeks*	**(ii)** • Work status
Kromer et al. (2014)		**(i)** *Age and gender were controlled for in the regression model
Mintken et al. (2010)		**(i)** • Age • Gender **(ii)** • Work status
Morrison et al. (1997)	**(i)** • Age: greater age; *worse pain, function, active ROM, strength and overall satisfaction via absolute UCLA score at average of 27 months*	**(i)** • Gender
Ogon et al. (2009)		**(i)** • Age • Gender **(ii)** • Occupation
Virta et al. (2009)		**(i)** • Age • Gender

Sindhu et al. (2012) and Tyler et al. (2010) did not explore demographic variables as potential prognostic factors.

The published studies in this area provided evidence of age and gender as the demographic factors most commonly predictive of outcome. Specifically, older age was associated with poorer absolute

outcome status at discharge (Deutscher et al. 2009), discharge / 12 weeks (Kennedy et al. 2006) and average of 27 months follow up (Morrison et al. 1997). In addition, younger age was predictive of greater improvement in disability via change in DASH score from baseline to discharge / 12 weeks (Kennedy et al. 2006). The direction of prediction from these studies is consistent and could be viewed as aligning with the theories of increasing age being associated with poorer tissue healing and/or enhanced ability to modify lifestyle and greater functional reserves amongst the young (Cecchi et al. 2014).

However it must be noted that a larger number of studies (n=8) did not identify age as a predictor. Of those studies identified as being of the highest level of quality (prospective cohort studies with >80% follow up), all that included age in their regression analysis found it to not be a predictor. Consequently the strength of evidence for age being a predictor is weakened.

Three studies (Kennedy et al. 2006; Deutscher et al. 2009; Chester et al. 2016) identified gender as a predictor of outcome and were consistent with female gender predicting poorer outcome. However these studies all measured absolute outcome at discharge rather than change from baseline. A larger number of studies (n=6) found gender not to be predictive of outcome, of which n=4 were prospective cohort studies and so in the higher end of the quality of prognostic study evidence hierarchy. Therefore, whilst age and gender were the most commonly identified demographic predictors of outcome, the strength of evidence is moderate.

The other generic demographic factors investigated were smoker status for which both studies that explored it (Deutscher et al. 2009; Chester et al. 2016) reported it to not be predictive of outcome. Chester et al. (2016) also considered BMI and the index of multiple deprivation but again neither were predictive of outcome. Finally Engebretsen et al. (2010) was the only study to consider educational attainment and identified low (defined as not having attending college or university) as predicting both worse pain and disability via absolute SPADI score and ‘not working’ status at 12 months post baseline.

With regards to work, recreational or functional task related variables, only three were identified as being predictive of outcome and none in more than one study. Work related variables were identified by Chester et al. (2016) as predicting worse absolute pain and disability follow up scores at 6 weeks and 6 months for those subjects not in employment due to redundancy, unemployment or disability, compared with being in employment or education. Kennedy et al. (2006) identified

presence of a workers compensation claim as predicting greater absolute disability at discharge / 12 weeks. Chester et al. (2016) was the only study to identify weekly exercise levels as predictive whilst no study identified functional tasks as being predictive. Indeed six studies did not find one or all of their work related variables as being predictors. Evidence of a consensus on work, recreational or functional task related variables predicting outcome in this topic area is therefore lacking.

In conclusion, a large number of different demographic related variables have been considered as potential prognostic factors in previous studies. However evidence for a consensus is lacking and there is only weak evidence for age and gender as predictors.

2.11.2 B. Clinical history

Table 2-3: B. Clinical history factors, segregated according to (i) Current condition related variables and (ii) Treatment related variables

Paper	Prognostic factor; *direction*	Not prognostic
	(i) Current condition related variables; (ii) Treatment related variables	**(i) Current condition related variables; (ii) Treatment related variables**
Bartolozzi et al. (1994)	**(i)** • Symptom duration: longer duration symptoms; *worse outcome via combination of UCLA score at discharge and perceived improvement at average of 20 months*	**(i)** • Side-related symptoms: dominant arm involvement • Reason for symptom onset
Chester et al. (2016)	**(i)** • Side-related symptoms: presence of pain in the opposite upper quadrant; *worse pain and disability via absolute SPADI score and worse disability via absolute Quick DASH score at 6 weeks post baseline* • Side-related symptoms: both shoulder affected or patient stated 'ambidextrous'; *worse disability via absolute Quick DASH score at 6 months post baseline* • Symptom duration: longer duration of symptoms; *worse pain and disability via absolute SPADI score*	**(i)** • Reason for symptom onset • Timing of onset of symptoms • History of previous shoulder pain • Paraesthesia in arm **(ii)** • Taking pain medication • Previously receiving treatment: physiotherapy helpful for previous shoulder problems

	and worse disability via absolute Quick DASH score at 6 months post baseline **(ii)** • Co-morbidities: additional health problems; *worse pain and disability via absolute SPADI score and worse disability via absolute Quick DASH score at 6 months post baseline*	
Conroy and Hayes (1998)		**(i)** • Side-related symptoms: dominant arm involvement • Symptom duration
Deutscher et al. (2009)	**(i)** • Symptom duration: chronic symptoms; *poorer functional status via absolute CAT score at discharge* **(ii)** • Taking antidepressant medication; *poorer functional status via absolute CAT score at discharge*	**(ii)** • Taking pain medication • Co-morbidities: number of co-morbidities
Engebretsen et al. (2010)	**(i)** • Recurrent shoulder pain: previous shoulder pain; *worse pain and disability via absolute SPADI score at 12 months post baseline*	**(i)** • Symptom duration • Associated neck pain • Side-related symptoms: dominant arm involvement **(ii)** • Previously receiving treatment: previous physiotherapy • Taking medication
Kennedy et al. (2006)	**(i)** • Symptom duration: shorter duration in symptoms; *greater improvement in disability via change from DASH score at baseline to discharge / 12 weeks* • Post-surgical case; *greater improvement in disability via change from DASH score at baseline to discharge / 12 weeks*	**(i)** • Recurrent shoulder pain • Reason for symptom onset **(ii)** • Co-morbidities: number of co-morbidities and number that limit activity • Taking pain medication
Kromer et al. (2014)	**(i)** • Symptom duration: longer duration of complaints; *less improvement in functional status via change in*	

	functional subscale of SPADI (SPADI-F) between baseline and 3 months post baseline	
Mintken et al. (2010)	**(i)** • Symptom duration: shorter duration of symptoms; *'improvement' via GROC score => 4 at discharge* **(ii)** • Taking pain medication: not taking pain medication; *'improvement' via GROC score => 4 at discharge*	**(i)** • Reason for symptom onset • Recurrent shoulder pain**: n**umber of previous episodes **(ii)** • Previously receiving treatment: treatment for previous episodes • Past medical history
Morrison et al. (1997)	**(i)** • Symptom duration: longer duration in symptoms; *worse score at average of 27 months*	**(i)** • Side-related symptoms: dominant arm involvement
Ogon et al. (2009)		**(i)** • Periods of professional disability • Reason for symptom onset • Symptom duration • Side-related symptoms: dominant arm involvement **(ii)** • Past medical history
Virta et al. (2009)		**(i)** • Symptom duration

Hung et al. (2010), Sindhu et al. (2012) and Tyler et al. (2010) did not explore clinical history variables as potential prognostic factors.

The clinical history variable most commonly identified as a predictor was duration of symptoms, whereby longer duration was associated with a worse outcome state (Morrison et al. 1997; Deutscher et al. 2009; Chester et al. 2016) and with worse change in status (Bartolozzi et al. 1994; Kromer et al. 2014). Similarly, shorter duration was predictive of greater improvement (Kennedy et al. 2006; Mintken et al. 2010) which together align with the concept that more established conditions are more stubborn to treatment (Collins et al. 2010). Four studies did not find it to be a predictor but only two of these were prospective cohort studies and so in the higher end of the quality of prognostic study evidence hierarchy. Therefore there is stronger evidence for duration of symptoms being a predictor than not; and where identified, the direction of prediction is consistent.

In relation to the other 'Current condition related variables', there was greater variability in findings. Chester et al. (2016) identified side related symptoms as being predictive of a worse outcome state, namely where patients had pain in the opposite upper quadrant and both shoulders affected or

patient stated ‘ambidextrous’. However, five studies found that symptoms defined relative to the limb side was not predictive, including high level of evidence studies (i.e. prospective cohort studies with >80% follow up). Similarly, previous shoulder pain episodes was found by Engebretsen et al. (2010) to be predictive of a worse outcome state but 3 other studies found it not to be predictive, including the larger and higher quality study by Chester et al. (2016). Of less ambiguity was the reason for onset of symptoms which was consistently identified as not being predictive in all five studies that investigated it. Similarly, the studies that considered neck pain or paraesthesia found them to not be predictive.

Treatment related variables included not taking pain medication or taking anti-depressant medication which Mintken et al. (2010) reported as predicting greater improvement and the larger study by Deutscher et al. (2009) reported as predicting a worse outcome state; respectively. However three studies found taking pain medication as not being predictive whilst Engebretsen et al. (2010) looked only at ‘use of medication’ and again found it to not be predictive. Similarly whilst Chester et al. (2016) found co-morbidities to predict a worse outcome state, four other studies found it to not be predictive. The three studies that looked at previous treatment for shoulder problems all found this variable to not be predictive.

It can therefore be concluded that a large range of clinical history related variables have been explored as potential prognostic factors in previous studies. For duration of symptoms there is moderate evidence for it predicting outcome and the direction of prediction is consistent. However for the remaining clinical history variables there is weak evidence of their predictive capacity.

2.11.3 C1. Patient reported measures: Pain

Table 2-4: C1. Patient reported measures: Pain

Paper	Prognostic factor; ***direction***	Not prognostic
Chester et al. (2016)	• Pain intensity: higher pain severity at rest; *worse pain and disability via absolute SPADI score and worse disability via absolute Quick DASH score at 6 weeks and 6 months post baseline*	
Engebretsen et al. (2010)		• Pain intensity

Kennedy et al. (2006)	• Pain intensity: higher pain intensity at baseline; *less improvement in disability via change from DASH score at baseline to discharge / 12 weeks*	
Kromer et al. (2014)		• Pain intensity
Mintken et al. (2010)		• Pain intensity

Bartolozzi et al. (1994), Conroy and Hayes (1998), Deutscher et al. (2009), Hung et al. (2010), Morrison et al. (1997), Ogon et al. (2009), Sindhu et al. (2012), Tyler et al. (2010) and Virta et al. (2009) did not explore pain variables as potential prognostic factors.

Three of the studies that explored pain as a potential prognostic factor found it not to be a predictor. However two of these were low in the hierarchy of prognostic evidence as they were RCT derived data and were at risk of over-fitting due to having less than 10 cases per variable. In contrast, the studies that found pain to be a predictor (Chester et al. 2016; Kennedy et al. 2006) were both prospective cohort studies and had 10 or more cases per variable, thereby providing higher quality evidence that was more likely to be replicated in a new sample. Specifically, Chester et al. (2016) identified higher pain at baseline as predicting a worse outcome state whilst the multiple regression equation coefficient reported by Kennedy et al. (2006) identified higher pain intensity predicting less improvement in status. Therefore there is moderate evidence of baseline pain levels being a predictor, with higher pain levels predicting worse outcome.

2.11.4 C2. Patient reported measures: Psychological symptoms

Table 2-5: C2. Patient reported measures: Psychological symptoms

Paper	Prognostic factor; *direction*	Not prognostic
Chester et al. (2016)	• Self-efficacy for pain: lower pain self-efficacy; *worse pain and disability via absolute SPADI score and worse disability via absolute Quick DASH score at 6 weeks and 6 months post baseline* • Patient expectation of 'slight improvement' rather than 'complete	• Anxiety and depression in the previous 7 days

	recovery' as 'a result of physiotherapy treatment; *worse pain and disability via absolute SPADI score and worse disability via absolute Quick DASH score at 6 weeks and 6 months post baseline* • Patient expectation of 'much improvement' rather than 'complete recovery' as 'a result of physiotherapy treatment'; *worse pain and disability via absolute SPADI score and worse disability via absolute Quick DASH score at 6 months post baseline* • Patient expectation of 'no change' rather than 'complete recovery' as 'a result of physiotherapy treatment'; *worse pain and disability via absolute SPADI score at 6 weeks post baseline*	
Engebretsen et al. (2010)	• General health status: poor, via baseline EQ-VAS; *'not working' status at 12 months post baseline*	• Emotional distress: Hopkins Symptom checklist • Self-efficacy for pain
Kennedy et al. (2006)		• Patient expectation • Mental Component Score (MCS) from the 36-Item Short-Form Health Survey (SF-36)
Kromer et al. (2014)		• Fear Avoidance Beliefs Questionnaire (physical activity subscale) • Pain catastrophising scale
Mintken et al. (2010)		• Patient expectation • Fear Avoidance Beliefs Questionnaire • Tampa Scale for Kinesiophobia
Sindhu et al. (2012)	• Fear Avoidance Beliefs Questionnaire: lower fear-avoidance beliefs; *greater functional improvement via change in CAT score between baseline and discharge*	

Bartolozzi et al. (1994), Conroy and Hayes (1998), Deutscher et al. (2009), Hung et al. (2010), Morrison et al. (1997), Ogon et al. (2009), Tyler et al. (2010) and Virta et al. (2009) did not explore psychological symptoms variables as potential prognostic factors.

Three of the six studies that explored state of mind or psychological symptoms as potential prognostic factors found one or more of them to be predictive of outcome. However there was substantial variation between studies in which variables were found to be predictive. Chester et al. (2016) reported three elements of patient expectation to be predictive of outcome, whilst neither Kennedy et al. (2006) nor Mintken et al. (2010) found patient expectation to be predictive. Chester et al. (2016) also found self-efficacy for pain to be a predictor whilst Engebretsen et al. (2010) did not.

Where more direct measures of state of mind or psychological symptoms were considered these were generally not found to be predictive of outcome. Indeed only Sindhu et al. (2012) found their state of mind variable of fear avoidance beliefs to be predictive of outcome and this was where only a single, dichotomised question was posed. However where fear avoidance beliefs were explored more fully as per Mintken et al. (2010) and Kromer et al. (2014) then this variable was not found to be predictive. Furthermore where informal (as per Chester et al. 2016) or formalised assessment (as per Engebretsen et al. 2010) of mental health was undertaken these were not found to be predictive. The strength of evidence pointing to state of mind or psychological symptoms being a predictor is therefore weak, although the wide variation in approaches used in previous studies compounds this further.

2.11.5 C3. Patient reported measures: Function / Disability

Table 2-6: C3. Patient reported measures: Function / Disability

Paper	Prognostic factor; *direction*	Not prognostic
Bartolozzi et al. (1994)	• Function / disability: more severe functional impairment as measured by the function element of the UCLA; *worse outcome via combination of UCLA score at discharge and perceived improvement at average of 20 months*	
Chester et al. (2016)	• Function / disability: higher baseline disability on SPADI; *worse pain and disability via absolute SPADI score at 6 weeks and 6 months post baseline*	

	• Function / disability: higher baseline disability on Quick DASH; *worse disability via absolute Quick DASH score at 6 weeks and 6 months post baseline*	
Deutscher et al. (2009)	• Function / disability: higher functional status via CAT at baseline; *better functional status via absolute CAT score at discharge*	
Engebretsen et al. (2010)	• Function / disability: higher pain and disability levels via baseline SPADI; *worse pain and disability via absolute SPADI score at 12 months post baseline*	
Hung et al. (2010)	• Function / disability: FLEX-SF score <41 (higher baseline disability); *'improvement' via GROC score => 4 at discharge*	
Kennedy et al. (2006)	• Function / disability: higher DASH (more disabled) at baseline; *greater disability via absolute DASH score at discharge / 12 weeks* • Physical Component Score (PCS) from the 36-Item Short-Form Health Survey (SF-36): worse physical health (SF-36: PCS) at baseline; *greater improvement in disability via change from DASH score at baseline to discharge / 12 weeks*	• Global rating of disability score
Kromer et al. (2014)	• Function / disability: higher disability levels as measured by the function element of the SPADI; *greater improvement in functional status via change in functional subscale of SPADI (SPADI-F) between baseline and 3 months post baseline*	
Mintken et al. (2010)		• Function / disability: SPADI

Conroy and Hayes (1998), Morrison et al. (1997), Ogon et al. (2009), Sindhu et al. (2012), Tyler et al. (2010), Virta et al. (2009) did not explore function / disability variables as potential prognostic factors.

Seven of the eight studies that explored function / disability found this variable category to be predictive of outcome and where outcome was defined in terms of absolute state, the direction of prediction was consistent. Specifically greater functional impairment / disability at baseline predicted a worse outcome state (Bartolozzi et al. 1994; Kennedy et al. 2006; Engebretsen et al. 2010; Chester et al. 2016) and higher functional status predicted better functional status at discharge (Deutscher et al. 2009).

Where the outcome was change of status, three studies (Kennedy et al. 2006; Hung et al. 2010; Kromer et al. 2014) found baseline function / disability to be predictive whilst only Mintken et al. (2010) did not. Both Hung et al. (2010) and Mintken et al. (2010) used the GROC although the very short study duration from Mintken et al. (2010) of approximately 1 week (from baseline to follow up) makes it difficult to meaningfully compare findings with the other studies in this area. With Kennedy et al. (2006), Hung et al. (2010) and Kromer et al. (2014) the direction of prediction was consistent, namely higher baseline disability was predictive of greater improvement.

Given the consistent finding that function / disability has been identified as a predictor, including in 4 prospective cohort studies, then the evidence to support this variable is strong. However the direction of prediction appears to be linked to how outcome is defined.

2.11.6 D1. Clinical measures: Strength

Table 2-7: D1. Clinical measures: Strength

Paper	Prognostic factor; *direction*	Not prognostic
Bartolozzi et al. (1994)		• Strength: shoulder girdle weakness
Chester et al. (2016)		• Strength: shoulder force (abduction, ER)
Hung et al. (2010)	• Strength: <27.4% body weight scapular protractor force; *'improvement' via GROC score => 4 at discharge*	• Strength: ER, IR and abduction
Kennedy et al. (2006)		• Strength: muscle wasting and muscle weakness

Mintken et al. (2010)		• Strength: Serratus Anterior, Middle trapezius, Lower trapezius, Rhomboid, Deltoid, External and internal shoulder rotator muscle strength

Conroy and Hayes (1998), Deutscher et al. (2009), Engebretsen et al. (2010), Kromer et al (2014), Morrison et al. (1997), Ogon et al. (2009), Sindhu et al. (2012), Tyler et al. (2010) and Virta et al. (2009) did not explore strength variables as potential prognostic factors.

The majority of studies (4 out of 5) found strength to not be predictive of outcome. In addition, a large number of individual strength variables were also not found to be predictive. It is noted that Hung et al. (2010) found scapular protractor force was predictive of improvement at discharge; however the sample size in this study was the smallest of those that considered strength as a potential predictor. As such the weight of evidence is that strength is not predictive of outcome.

2.11.7 D2. Clinical measures: ROM

Table 2-8: D2. Clinical measures: ROM

Paper	Prognostic factor; *direction*	Not prognostic
Bartolozzi et al. (1994)		• ROM: active abduction and flexion ROM
Chester et al. (2016)	• ROM: reduced range of active shoulder abduction; *worse pain and disability via absolute SPADI score at 6 weeks post baseline* • ROM: increasing difference between the range of active and passive shoulder abduction; *worse pain and disability via absolute SPADI score at 6 months post baseline*	• ROM: AROM (flexion) • ROM: PROM (flexion, abduction, ER)
Engebretsen et al. (2010)		• ROM: AROM – HBB • ROM: flexion on the affected side
Hung et al. (2010)		• ROM: PROM IR, ER • ROM: posterior shoulder tightness (PST)
Kennedy et al. (2006)		• ROM: active and passive ROM

Mintken et al. (2010)	• ROM: <127° of pain-free shoulder flexion; *'improvement' via GROC score => 4 at discharge* • ROM: <53° passive shoulder IR at 90° abduction; *'improvement' via GROC score => 4 at discharge*	• ROM: Passive shoulder abduction • ROM: Passive shoulder ER at 90° abduction • ROM: Battery of 3 functional tests
Tyler et al. (2010)	• ROM: Improvement in PST at discharge; *symptom free via Simple Shoulder Test at discharge*	• ROM: Improvement in passive ER ROM • ROM: Improvement in GIRD

Conroy and Hayes (1998), Deutscher et al. (2009), Kromer et al (2014), Morrison et al. (1997), Ogon et al. (2009), Sindhu et al. (2012) and Virta et al. (2009) did not explore ROM variables as potential prognostic factors.

ROM variables were found to be predictive of outcome in three of the seven studies that explored this variable. However of these 3 studies the ROM variables were relatively disparate, ranging from static measures of soft tissue tightness (Tyler et al. 2010) to passive ROM and pain free active ROM (Mintken et al. 2010) to the difference between active and passive ROM (Chester et al. 2016). This makes it difficult to establish a consensus as to what aspects of ROM might be predictive. The large number of ROM variables not found to be predictive in all seven studies makes any such consensus even more tenuous.

2.11.8 D3. Clinical measures: Scapular movement and control

Table 2-9: D3. Clinical measures: Scapular movement and control

Paper	Prognostic factor; ***direction***	Not prognostic
Chester et al. (2016)	• Scapular movement and control: No change compared to change in shoulder pain/ range during manual facilitation of the scapula around the chest wall during arm elevation; *worse pain and disability via absolute SPADI score and worse disability via absolute Quick DASH score at 6 months post baseline*	
Hung et al. (2010)	• Scapular movement and control: <0.7° scapular IR at 30° elevation during descending phase in the unloaded condition; *'improvement' via GROC score => 4 at discharge*	• Scapular movement and control: Scapular kinematics (UR, IR and scapular tilt) at 30°, 60°, 90°, 120° elevation during both the ascending

		and descending phases in loaded and unloaded conditions
Mintken et al. (2010)		• Scapular movement and control: Lateral slide test and scapula index • Scapular movement and control: Qualitative assessment of scapular function

Bartolozzi et al. (1994), Conroy and Hayes (1998), Deutscher et al. (2009), Engebretsen et al. (2010), Kennedy et al. (2006), Kromer et al (2014), Morrison et al. (1997), Ogon et al. (2009), Sindhu et al. (2012), Tyler et al. (2010) and Virta et al. (2009) did not explore scapular movement and control variables as potential prognostic factors.

The potential for scapular movement and control variables to predict outcome was only explored in three previous studies, two of whom identified it as being a predictor. However whilst Chester et al. (2016) explored this variable using essentially a symptom modification test, Hung et al. (2010) was measuring detailed movements using a 3-dimensional movement analysis system. Of note here is that in the Hung et al. (2010) study only one of the three movements at one of the four positions, during one of the two phases and under one of the two loaded trials provided a predictive variable. Along with the absence of prediction from the static measures and qualitative assessment used by Mintken et al. (2010), a lack of consensus is seen in the previous literature.

2.11.9 E1. Structural pathology via imaging

Table 2-10: E1. Structural pathology via imaging

Paper	Prognostic factor; *direction*	Not prognostic
Bartolozzi et al. (1994)	• Rotator cuff pathology: (moderate or large rotator cuff tears) on MRI; *worse outcome via combination of UCLA score at discharge and perceived improvement at average of 20 months*	
Morrison et al. (1997)	• Acromion morphology: (type 1) on X-ray; *better outcome at average of 27 months*	
Ogon et al. (2009)	• Calcific deposits: Radiographic classification of the calcific deposits and the sonographic penetration of the deposits; *'success of non-operative therapy' via no progression to advanced therapeutic measures*	• Calcific deposits: Multifocal distribution of calcific deposits • Calcific deposits: French Society of Arthroscopy classification of calcific deposits

	after a minimum of 6 months non-operative treatment (including minimum of 3 months treatment at the study location) • Calcific deposits: Bilateral calcific deposits, location and volume of the deposits on X-ray; *'failure of non-operative therapy' via progression to advanced therapeutic measures due to persistence of symptomatic calcific tendonitis after a minimum of 6 months non-operative treatment (including minimum of 3 months treatment at the study location)*	

Chester et al (2016), Conroy and Hayes (1998), Deutscher et al. (2009), Engebretsen et al. (2010), Hung et al. (2010), Kennedy et al. (2006), Kromer et al (2014), Mintken et al. (2010), Sindhu et al. (2012), Tyler et al. (2010) and Virta et al. (2009) did not explore structural pathology via imaging variables as potential prognostic factors.

The potential for structural pathology via imaging variables to predict outcome was only explored in three previous studies (Bartolozzi et al. 1994; Morrison et al. 1997; Ogon et al. 2009), all of whom identified it as being a predictor. Each study used a different measurement tool and identified different structures or tissues as being predictive. The presence of these different structures or tissues was predictive of both improved outcome (Morrison et al. 1997; Ogon et al. 2009) and worse outcome (Bartolozzi et al. 1994; Ogon et al. 2009). In addition, Ogon et al. (2009) found multiple structural pathology via imaging variables to not be predictive. It should also be noted that two (Bartolozzi et al. 1994; Morrison et al. 1997) of the studies were retrospective cohort studies and so provide the lowest quality of prognostic evidence. It can therefore be concluded that there is only moderate evidence for structural pathology via imaging variables to be predictive. There is also a lack of consensus as to what tools and structures / tissues generate predictive variables or in what direction they predict.

2.11.10 E2. Structural pathology via orthopaedic tests

Table 2-11: E2. Structural pathology via orthopaedic tests

Paper	Prognostic factor; *direction*	Not prognostic
Chester et al. (2016)		• Orthopaedic tests: External lag sign

Mintken et al. (2010)	• Orthopaedic tests: Negative Neer's sign; *'improvement' via GROC score => 4 at discharge*	• Orthopaedic tests: 12 other orthopaedic tests, including the Hawkins-Kennedy impingement test, the empty can and full can test and the drop sign

Bartolozzi et al. (1994), Conroy and Hayes (1998), Deutscher et al. (2009), Engebretsen et al. (2010), Hung et al. (2010), Kennedy et al. (2006), Kromer et al (2014), Morrison et al. (1997), Ogon et al. (2009), Sindhu et al. (2012), Tyler et al. (2010) and Virta et al. (2009) did not explore orthopaedic tests variables as potential prognostic factors.

Only two previous studies investigated the potential of structural pathology via orthopaedic tests to predict outcome. Of these only Mintken et al. (2010) found them to be predictive and this was only the Neer's sign out of a battery of 13 tests. However, as the Mintken et al. (2010) study undertook treatment and follow up over a time period of only a few days, the clinical utility of their findings is very weak. As such there is weak evidence for orthopaedic tests being predictive of outcome.

2.11.11 Summary of individual prognostic variables

Using a Cochrane style 'traffic light' summary, the evidence for the variables being predictive from the previous literature can be coded as follows:

Age & gender

Duration of symptoms

Pain

Psychological state

Function

Strength

ROM

Scapula movement control

Structural pathology via imaging

Structural pathology via orthopaedic tests

Where green denotes moderate or strong evidence, yellow denotes weak evidence or lacking in consensus and red denotes evidence that it is not predictive. However the utility of these judgements in informing the objectives of this study (including how to identify the potential prognostic factors to investigate) is limited by (i) the short-comings of attempting to synthesise evidence regarding individual prognostic factors rather than actual prognostic models (Braun et al. 2016), (ii) the PROBAST guided quality appraisal which identified all studies as being at high risk of bias, the majority where there was high level concern regarding applicability to the research

question and no studies identified as being useable in the context of the intended context and target population which in turn limits the ability to synthesise the evidence.

It can however be concluded that potential prognostic factors from a wide range of categories have been identified in each of the 10 categories (A, B, C1, C2, C3, D1, D2, D3, E1 and E2) – and these each have conceptual relevance – around which the literature review on prognostic factors was structured. As such this provides a rationale and a context for examining these variables in this thesis.

2.11.12 Multivariate predictors

As noted previously, multivariate regression techniques enable examination of the interaction of multiple variables in predicting a particular outcome (Hemingway et al. 2009). Where studies used end state as their outcome variable then the number of variables comprising their predictive model ranged from three (Engebretsen et al. 2010) to four (Kennedy et al. 2006) and five (Deutscher et al. 2009). Where multiple outcome variables and time points were used, then between three (for Quick DASH at 6 weeks) and six (for SPADI at 6 months) comprised the model (Chester et al. 2016). Where change in state was the outcome variable then the number of variables comprising predictive models ranged from two (Kromer et al. 2014) to three (Hung et al. 2010) and five (Kennedy et al. 2006). In the previously published literature, wide variation was therefore seen in the number of variables comprising the final predictive models.

2.12 Overall summary

This review of the literature has highlighted the wide range of potential prognostic categories and variables that have collectively been explored by previous studies in this area. This is perhaps unsurprising given the uncertain nature of SIS/RCTendinopathy in regard to its aetiology, signs and symptoms, how best to define it and the conceptual frameworks within which it can be contextualised. In light of this persistent uncertainty then there is merit in incorporating the full range of potential prognostic categories within a single study. Such an approach aligns well with the complex, multi-factorial intervention of physiotherapy. Recognising the historically poor standard of prognostic studies in both musculoskeletal and non-musculoskeletal research areas, an emphasis on high quality study design is imperative.

Identification of the patient population was explored early on in this literature review, including the limitations of relying on orthopaedic tests for diagnosis due to their limited sensitivity and

specificity. Yet as the most commonly diagnosed shoulder pathology, a clinically and conceptually meaningful approach to identifying an SIS/RCTendinopathy patient sample is essential. As such, the exclusion of discrete shoulder pathologies provides one mechanism to identify a patient sample, combined with a clinical presentation suggestive of the target pathology. Another aspect of identifying a patient sample is to identify where in the patient care pathway should subjects be recruited from. Ideally this should be based upon an established, evidence-based, optimum care pathway. However this is currently lacking for patients with SIS/RCTendinopathy and so the use of a meaningful and well-defined section of the care pathway provides a constructive compromise.

This literature review has highlighted a number of study design limitations with previous studies in this area. Given the emphasis in the prognostic methodology literature on high standards, the use of a prospective cohort study with >80% follow up is essential. Further failings in the previous literature include not examining the characteristics of those who did and did not consent as a potential source of selection bias. Comparison of baseline characteristics between those who were available for follow up and those lost to follow up was also rarely performed. As a potential source of bias within the final analysis groups it is essential that such analysis is performed, as is blinding of assessors to the individual patient's outcome until all other data collection and processing is completed.

By including the full range of potential prognostic categories within a single study, there is merit in collecting (A) demographics and (B) clinical history data. However, given the large number of potential variables within these categories, a judicious decision should be made as to which to include in the prognostic model. For (A) demographics, age and gender are two likely candidates whilst for (B) clinical history data, duration of symptoms is the variable which the literature indicates has the most predictive potential. Nonetheless the collecting of a wide range of (A) demographics and (B) clinical history variables provides a mechanism for comprehensively defining the sample and is therefore advocated.

In relation to (C) patient reported measures, there is moderate evidence of (C1) pain measured by the VAS as predicting outcome. However the wide range of references for quantifying it (e.g. at rest, on movement, etc) and timescales for recall of pain levels that have been used in the previous literature undermine any such consensus. As such the use of a well reasoned approach to identifying the pain variable to be included in the final prognostic model should be made.

In relation to (C2) psychological symptoms, the wider musculoskeletal literature provides a strong rationale for including this variable type. However the prognostic literature provides a less unanimous picture of its predictive ability and this is compounded by both the complexity of psychological symptoms and the myriad measurement tools that have previously been used to measure it. As such, the use of an established tool for capturing the wide range of psychological symptoms in a clinically interpretable manner is advocated.

When considering patient reported measures of (C3) function and disability, the previous literature provides unanimous evidence of the predictive value of this variable type. However of greater challenge is which tool to use given the wide range of outcome scores available in the published literature. The frequent use in previous studies of the same outcome score as both a candidate variable and the definer of outcome is a notable limitation of such studies and must be addressed in any future study.

The conceptual relevance of (D) clinical measures such as (D1) strength, (D2) ROM and (D3) scapular movement and control are represented in both the path-anatomical and the International Classification of Functioning, Disability and Health derived models of SIS/RCTendinopathy. However this contrasts with limited evidence from previous studies that strength is predictive of outcome; whilst establishing a consensus as to which aspects of ROM might be predictive is challenging given the wide range of approaches previously used. The use therefore of well reasoned measures of strength and ROM along with formal testing of the reliability of the tools in the context within which they are used should be undertaken. In the case of (D1) strength, consideration of this as a potential prognostic variable – in spite of the weight of evidence being that strength is not predictive of outcome – is rationalised by such evidence being only one factor in determining which variables to consider, combined with the clear conceptual relevance of this variable.

Given the clearly recognised challenges of measuring scapular movement and control in the clinical setting, issues around reproducibility of any tool pose a particular issue given the emphasis within the prognostic methodology literature on using robust measurement approaches. Therefore alongside a judicious choice of measurement tool, the formal testing of the reproducibility of any tool should be undertaken.

The exploration of (E) structural pathology is a less commonly used approach in the previous literature, including the use of (E1) imaging modalities. The ability for diagnostic ultrasound to

provide high resolution, clinic-based imaging of a wide range of potentially relevant shoulder structures makes it an attractive option. However given the highly technical nature of performing and interpreting diagnostic ultrasound, any such inclusion of the modality must be premised upon the use of an appropriately trained operator. Furthermore, given the subjective nature of imaging interpretation, formal reproducibility data should be provided. The exploration of (E2) orthopaedic test results as a potential predictor variable has rarely been undertaken in the literature. Given their limited diagnostic value, inclusion of orthopaedic tests in a prognostic model reflects one of the notable advantages of prognostic research, namely that it is not constrained by diagnosis *per se*.

Due to the complex, multi-factorial nature of physiotherapy, the use by many previous studies of a narrowly defined approach to treatment makes their findings of limited clinical relevance. As such there is a strong case for using a pragmatic approach to treatment. However the challenges of recording and quantifying the intervention are substantial. When considering what factors at the outset can predict outcome then the omitting of treatment related data from prognostic modelling is appropriate. However awareness of the content, duration and frequency of treatment – along with response to the intervention and levels of non-completion of treatment – allow a study to be set in the context of routine care.

Alongside the process by which a prognostic model is constructed, the mechanism by which outcome is defined and measured has a substantial impact on the subsequent findings. The consistent use in recent studies of patient reported measures enables the capturing of the wide impact of SIS/RCTendinopathy and is therefore advocated. As however is the importance of ensuring no absolute overlap with any measurement tools used for capturing potential prognostic factors. In the published literature there has been wide variation in the timescales used for follow up and this limits both the clinical relevance of their results and the ability to pool findings. An optimal approach is to align follow up with the actual completion of active treatment alongside use of a standardised follow up period post-discharge. Changes in patient reported measures from baseline provides a highly clinically relevant mechanism for processing the dependent variable.

In considering the full range of potential prognostic categories within a single study, one notable issue is how to address the risk of over-fitting due to the inclusion of too many variables relative to the sample size. Several of the previously published studies in this area were identified as being at risk of this, thereby undermining the potential value of their evidence. One mechanism to overcome this is the defining *a priori* of a reasoned mechanism to select variables to be considered in a

prognostic model and to align the number of these to the actual recruited sample size. Due to the inevitable compromises that such an approach requires, placing a prognostic research study within a method framework whereby the nature and interpretation of studies sits within stages of prognostic research development (Hemingway et al. 2013) has merit. As such the emphasis on methodological rigour and exploration of factors across a range of types of variables would allow a study of this type to sit at the exploratory end of the prognostic framework (Kent et al. 2010), thereby providing a foundation for informing hypothesis-testing in a new sample.

This review of the literature will therefore inform the design and interpretation of data from a prospective study to explore prognostic indicators of successful rehabilitation outcome in patients with SIS/RCTendinopathy.

3 AIMS, OBJECTIVES AND HYPOTHESES

3.1 Aim

The aim of this study is to identify baseline factors that predict outcome in SIS/RCTendinopathy patients following physiotherapy treatment.

3.2 Research question

Which baseline variable or combination of baseline variables predict outcome following physiotherapy treatment in a cohort of SIS/RCTendinopathy patients?

3.3 Null Hypotheses

A.(i) Demographics

In a multivariate analysis, demographics will not be a predictor of outcome between baseline and discharge in SIS/RCTendinopathy patients treated with physiotherapy.

B.(i) Clinical history related

In a multivariate analysis, clinical history related variable will not be a predictor of outcome between baseline and discharge in SIS/RCTendinopathy patients treated with physiotherapy.

C.(i) Patient reported measures

In a multivariate analysis, baseline patient reported measures will not be a predictor of outcome between baseline and discharge in SIS/RCTendinopathy patients treated with physiotherapy.

D.(i) Clinical measures

In a multivariate analysis, baseline clinical measures will not be a predictor of outcome between baseline and discharge in SIS/RCTendinopathy patients treated with physiotherapy.

E.(i) Structural pathology

In a multivariate analysis, baseline structural pathology will not be a predictor of outcome between baseline and discharge in SIS/RCTendinopathy patients treated with physiotherapy.

F.(i) Multivariate prognosis

In a multivariate analysis, a combination of 2 or more of A.(i) demographics, B.(i) clinical history related, C.(i) baseline patient reported measures, D.(i) baseline clinical measures and E.(i) baseline

structural pathology will not predict outcome between baseline and discharge in SIS/RCTendinopathy patients treated with physiotherapy.

A.(ii) Demographics

In a multivariate analysis, demographics will not be a predictor of outcome between baseline and 3 months post-discharge in SIS/RCTendinopathy patients treated with physiotherapy.

B.(ii) Clinical history related

In a multivariate analysis, clinical history related variable will not be a predictor of outcome between baseline and 3 months post-discharge in SIS/RCTendinopathy patients treated with physiotherapy.

C.(ii) Patient reported measures

In a multivariate analysis, baseline patient reported measures will not be a predictor of outcome between baseline and 3 months post-discharge in SIS/RCTendinopathy patients treated with physiotherapy.

D.(ii) Clinical measures

In a multivariate analysis, baseline clinical measures will not be a predictor of outcome between baseline and 3 months post-discharge in SIS/RCTendinopathy patients treated with physiotherapy.

E.(ii) Structural pathology

In a multivariate analysis, baseline structural pathology will not be a predictor of outcome between baseline and 3 months post-discharge in SIS/RCTendinopathy patients treated with physiotherapy.

F.(ii) Multivariate prognosis

In a multivariate analysis, a combination of 2 or more of A.(i) demographics, B.(i) clinical history related, C.(i) baseline patient reported measures, D.(i) baseline clinical measures and E.(i) baseline structural pathology will not predict outcome between baseline and 3 months post-discharge in SIS/RCTendinopathy patients treated with physiotherapy.

3.4 Experimental Hypotheses

A.(i) Demographics

In a multivariate analysis, demographics will be a predictor of outcome between baseline and discharge in SIS/RCTendinopathy patients treated with physiotherapy.

B.(i) Clinical history related

In a multivariate analysis, clinical history related will be a predictor of outcome between baseline and discharge in SIS/RCTendinopathy patients treated with physiotherapy.

C.(i) Patient reported measures

In a multivariate analysis, baseline patient reported measures will be a predictor of outcome between baseline and discharge in SIS/RCTendinopathy patients treated with physiotherapy.

D.(i) Clinical measures

In a multivariate analysis, baseline clinical measures will be a predictor of outcome between baseline and discharge in SIS/RCTendinopathy patients treated with physiotherapy.

E.(i) Structural pathology

In a multivariate analysis, baseline structural pathology will be a predictor of outcome between baseline and discharge in SIS/RCTendinopathy patients treated with physiotherapy.

F.(i) Multivariate prognosis

In a multivariate analysis, a combination of 2 or more of A.(i) demographics, B.(i) clinical history related, C.(i) baseline patient reported measures, D.(i) baseline clinical measures and E.(i) baseline structural pathology will predict outcome between baseline and discharge in SIS/RCTendinopathy patients treated with physiotherapy.

A.(ii) Demographics

In a multivariate analysis, demographics will be a predictor of outcome between baseline and 3 months post-discharge in SIS/RCTendinopathy patients treated with physiotherapy.

B.(ii) Clinical history related

In a multivariate analysis, clinical history related will be a predictor of outcome between baseline and 3 months post-discharge in SIS/RCTendinopathy patients treated with physiotherapy.

C.(ii) Patient reported measures

In a multivariate analysis, baseline patient reported measures will be a predictor of outcome between baseline and 3 months post-discharge in SIS/RCTendinopathy patients treated with physiotherapy.

D.(ii) Clinical measures

In a multivariate analysis, baseline clinical measures will be a predictor of outcome between baseline and 3 months post-discharge in SIS/RCTendinopathy patients treated with physiotherapy.

E.(ii) Structural pathology

In a multivariate analysis, baseline structural pathology will be a predictor of outcome between baseline and 3 months post-discharge in SIS/RCTendinopathy patients treated with physiotherapy.

F.(ii) Multivariate prognosis

In a multivariate analysis, a combination of 2 or more of A.(i) demographics, B.(i) clinical history related, C.(i) baseline patient reported measures, D.(i) baseline clinical measures and E.(i) baseline structural pathology will predict outcome between baseline and 3 months post-discharge in SIS/RCTendinopathy patients treated with physiotherapy.

4 PRE-METHODS

The primary focus of this thesis is a prospective cohort study of SIS/RCTendinopathy patients to identify what variable or combination of variables predict outcome. Specifically these variables will be A. demographics, B. clinical history related, C. patient reported measures (C1. pain; C2. psychological symptoms; C3. function/disability), D. clinical measures (D1. strength; D2. ROM; D3. scapular movement and control) and E. structural pathology (via E1. imaging; E2. orthopaedic tests). For reasons of consistency, all methods by which the cohort study was undertaken will be detailed in the methods chapter.

However numerous preliminary studies were performed to inform the methods and analysis of the prognostic cohort study data and these comprise the current chapter. As per the GRRAS checklist (Kottner et al. 2011) they are summarised in appendix IV. Specifically, intra-rater reliability was required to be assessed for clinical measures: D1. strength and D2. ROM. Inter-rater reliability was assessed for D3. clinical measures: scapular movement and control and E1. structural pathology via imaging. Furthermore, formal post-graduate clinical training was undertaken for E1. structural pathology via ultrasound imaging. In addition a bespoke tool was developed for recording the physiotherapy delivered which comprised the pragmatic treatment.

4.1 D Clinical measures

4.1.1 D1. Clinical measures: Strength

In this prognostic cohort study, isometric strength ('make' test) was measured as a potential prognostic variable. A handheld dynamometer (Commander Muscle Tester, JTech Medical, USA) was used by Michael Smith (MS) to quantify the force generated by each subject. Full details of how strength was measured will be provided in the method section (page 126). However an intra-rater reliability study was conducted to determine how reliable the data collection method was in the prognostic cohort study.

4.1.1.1 Strength intra-rater reliability study

The handheld dynamometer has high levels of force accuracy as reported by the manufacturer (99% force accuracy over the entire test pad area; http://www.jtechmedical.com/phocadownload/manuals/MN084-CommanderMuscleTesterManual.pdf; accessed 31/10/2016); however this was not independently tested. As a single rater (MS) performed the prognostic cohort study strength measurements, it was

deemed necessary to assess the absolute accuracy of the strength measures taken in the same test positions using the same measurement tool as for the prognostic cohort study (Schrama et al. 2014).

In order to minimise the within subject, within-side variability of the force generated by the shoulders to be tested, the reproducibility study was undertaken on healthy subjects as they were deemed more able to repeatedly generate a consistent force (Awatani et al. 2016). The reproducibility study was undertaken on a convenience sample of 12 healthy subjects (4 male, 8 female; 1 left handed, 11 right handed). The demographics and OSS scores for these subjects are shown in the table below and the raw data is in appendix 4. The corresponding data collection form can also be seen in appendix 4.

Table 4-1: Demographics of strength intra-rater reliability study subjects

Variable	Measure of variability			
	Mean	SD	Minimum	Maximum
Age (years)	38.6	12.7	21.0	67.0
Height (m)	1.73	0.11	1.57	1.91
Weight (kg)	75.0	10.8	63.3	93.6
BMI (kg/m^2)	25.2	2.8	21.9	31.8
OSS	47.5	1.0	45	48

Key: Kg = kilograms, m = metres, BMI = body mass index, OSS = Oxford shoulder score, SD = standard deviation

It can be seen from the above table that the mean OSS for the reproducibility subjects was 47.5 with a range of 45 to 48. This is well within the range reported in the literature for asymptomatic subjects (Younis et al. 2011).

4.1.1.1.1 Procedure

The strength measurements in the reproducibility study were performed in the same test positions, with the same instructions to the subject and in the same order as in the prognostic cohort study. The distance from the axis of rotation to the point of force application was measured in metres using a measuring tape (Awatani et al. 2016). Three strength measures were taken on each limb side for each contraction type. The order was internal rotation (IR; right then left), external rotation (ER; right then left) and finally elevation in the scapular plane (Scaption; right then left). No shoulder pain was reported by any reproducibility study subjects during testing.

4.1.1.1.2 Data processing

The rotational moments (Newton metres) were calculated by multiplying the force generated (Newtons) during IR, ER and Scaption multiplied by the relevant moment arm (metres) (Schrama et al. 2014).

4.1.1.1.3 Data analysis

The intra-rater reliability study was concerned with the absolute agreement between measures taken by the same, single rater (MS) who took all the measures both in the intra-rater reliability study and in the prognostic cohort study. Thus, intraclass correlation coefficient ($ICC_{3,1}$) (Shrout and Fleiss 1979) was calculated in Statistical Package for the Social Sciences (SPSS version 20; IBM) where the model was 'two way mixed', type was 'absolute agreement' and the 'single measures' variable reported (Portney and Watkins 2007).

4.1.1.1.4 Results

The raw data of the strength readings for each trial, each contraction type and each limb side (IR, ER, Scaption; left, right) is in appendix 4 and the measures of variability for the contraction type and limb side combination is shown in the below table.

Table 4-2: Descriptive data of strength readings (Nm) for the strength measures and limb side combination trials from the intra-rater reliability study

Variable	Measure of variability			
	Mean	SD	Minimum	Maximum
IR Left	21.3	7.7	9.7	35.0
IR Right	20.8	8.5	8.7	37.3
ER Left	17.4	5.8	9.7	31.9
ER Right	18.0	6.3	9.7	29.4
Scaption Left	25.0	7.0	11.9	36.6
Scaption Right	25.1	7.9	11.9	37.3

Key: Nm = Newton metres, SD = standard deviation

Table 4-3: Intraclass Correlation Coefficient (ICC(3,1)) results for strength intra-rater reliability study

Variable	Intraclass Correlation Coefficient		
	$ICC_{(3,1)}$	95% CI (Lower bound)	95% CI (Upper bound)
IR Left	0.978	0.944	0.993
IR Right	0.973	0.932	0.992
ER Left	0.896	0.753	0.966
ER Right	0.966	0.915	0.989
Scaption Left	0.863	0.680	0.955
Scaption Right	0.881	0.685	0.962

Key: CI = confidence interval

4.1.1.1.5 Interpretation

The intra-rater reliability findings for the left and right side for IR, ER and Scaption are all in the very good range (Landis and Koch 1977; Altman 1991). Therefore the data collection method can be considered to be reliable and so the tool could be used in the same format in the prognostic cohort study.

4.1.2 D2. Clinical measures: ROM

In this prognostic cohort study, active planar ROM was assessed as a potential prognostic variable. A digital inclinometer (DigiPas DWL-180s, JSB Tech Pte Ltd, Singapore) was used to quantify the angular ROM achieved by each patient. Full details of how ROM was measured will be provided in the method section (page 128). However an intra-rater reliability study was conducted to determine how reliable the data collection method was in the prognostic cohort study.

4.1.2.1 ROM intra-rater reliability study

The digital inclinometer has high levels of angular accuracy as reported by the manufacturer (±0.05° at 0° and 90°; and ±0.2° at 1° to 89° (http://www.digipas.co.uk/products/digital-level-module/dwl-180.php; accessed 02/11/2016); however this was not independently tested. A single rater (MS) performed the prognostic cohort study ROM measurement and so it was deemed necessary to assess the intra-rater reliability of this tool in the context within which it was to be used in the prognostic cohort study. This involved assessing the absolute accuracy of the procedure of placing the inclinometer on the subject's arm and taking a reading.

The reproducibility study was undertaken on the same convenience sample of 12 healthy subjects as the strength reproducibility study. In order to remove the potential variable of the subjects moving their arm between measures, the reproducibility study was undertaken on healthy subjects as they were more likely to be able to maintain a consistent arm position.

4.1.2.1.1 Procedure

In order to replicate the likely ROM positions which the prognostic cohort study subjects would achieve, the ROM intra-rater reliability study subjects were measured in various sections of the available range. Vertical poles were used to standardise the plane of movement and a marker was placed on each pole (left and right) to which the subjects moved their arms. One was at a point that was <90° elevation and the other at >90° elevation. The order of this in relation to arm side (left / right) was counterbalanced within the sample.

The subjects were instructed to elevate their arms bilaterally to the respective markers. MS placed the digital inclinometer on the mid-point of the lateral aspect of the subject's upper arm, parallel to the humerus. A reading was taken, the digital inclinometer was removed from the subject's upper arm and the reading recorded on a sheet of paper. This procedure was repeated a further 2 times. The above procedure was then repeated for the other arm and recorded on the data collection form in appendix 4.

4.1.2.1.2 Data processing

ROM readings were converted from the reading on the unit (where 0° represents horizontal) to the absolute ROM attained by adding the reading to 90 (for above horizontal positions) and subtracting the reading from 90 (for below horizontal positions).

4.1.2.1.3 Data analysis

The intra-rater reliability study was concerned with the absolute agreement between measures taken by the same, single rater (MS) who took all the measures both in the intra-rater reliability study and in the prognostic cohort study. Thus, intraclass correlation coefficient ($ICC_{3,1}$) (Shrout and Fleiss 1979) was calculated in SPSS where the model was 'two way mixed', type was 'absolute agreement' and the 'single measures' variable reported (Portney and Watkins 2007).

4.1.2.1.4 Results

The raw data of the ROM readings for each trial and position (<90° elevation and >90° elevation) is in appendix 4 and the measures of variability for the <90° and >90° elevation trials is shown in the below table.

Table 4-4: Descriptive data of ROM readings (/°) for the <90° and >90° elevation trials from the intra-rater reliability study

Variable	Measure of variability			
	Mean	SD	Minimum	Maximum
<90° elevation	43.7	9.6	26.6	62.3
>90° elevation	135.7	7.2	119.5	151

Key: SD = standard deviation

Table 4-5: Intraclass Correlation Coefficient (ICC(3,1)) results for ROM intra-rater reliability study

Variable	Intraclass Correlation Coefficient		
	$ICC_{(3,1)}$	95% CI (Lower bound)	95% CI (Upper bound)
<90° elevation	0.980	0.949	0.994
>90° elevation	0.922	0.813	0.975

Key: CI = confidence interval

4.1.2.1.5 Interpretation

The intra-rater reliability findings for both <90° elevation and >90° elevation are both in the very good range (Landis and Koch 1977; Altman 1991). Therefore the data collection method can be considered to be reliable and so the tool could be used in the same format in the prognostic cohort study.

4.1.3 D3. Clinical measures: Scapular movement and control

In this prognostic cohort study, scapular movement and control was assessed as a potential prognostic variable. The scapular dyskinesis test (SDT) described by McClure et al. (2009) was used to categorise the presence and extent of any altered scapular movement and control. Full details of how scapular movement and control was measured will be provided in the method section (pg 129).

However as a subjective rating system it was deemed necessary for MS to undertake an inter-rater reliability study.

4.1.3.1 Scapular dyskinesis test inter-rater reliability study

Details of the validity (Tate et al. 2009) and inter-rater reliability of the SDT (McClure et al. 2009) have been published previously. However the SDT was developed on overhead athletes whilst the current study is concerned with NHS patients who were a non-athletic population. The likely greater heterogeneity of NHS patients in terms of body habitus and range of movement at the shoulder complex meant that establishing the level of agreement between the MS and another experienced musculoskeletal clinician using a sample of such patients was necessary.

As with the main cohort data, rating of scapular movement and control was performed on video recordings of the patients' movement patterns. To ensure direct transferability of the inter-rater reliability findings to the main cohort, the reliability study was undertaken on a subset of patients from the main cohort and level of agreement in rating the symptomatic side was established with one of the treating musculoskeletal clinicians in the study. This was Kevin Nicholas (KN) who worked at the Whitchurch hospital (WHI) site and had 10 years' experience working in musculoskeletal outpatients. In order to remove the potential confounding factor of being involved with the treatment of a subject, the sub-set of patients were ones that were treated at a different clinical site to where KN worked.

As with the published SDT reproducibility study (McClure et al. 2009), KN and MS undertook standardised training, which comprised the operational definitions, rating scale and photographic examples provided in the original publication. As the SDT is based upon interpretation of descriptive terminology, it was considered essential that the training and familiarisation process included applying the SDT to footage of scapular movement. Twenty videos of scapular movement patterns from patients with SIS/RCTendinopathy from both loaded (i.e. holding a weight in each hand) and un-loaded (i.e. not holding a weight) trials which demonstrated a spectrum of scapular movement patterns were therefore selected from the prognostic cohort. KN and MS viewed these videos and had the opportunity to discuss the interpretation and application of the descriptive terminology and recording process. This mirrored a typical, peer-supported clinical training environment (Baertschi et al. 2013).

Following the training period, KN and MS each had a copy of the inter-rater reliability study videos and data collection forms. These inter-rater reliability study videos were of a different 30 subjects which were not used in the training phase. Each assessor independently graded each of the videos with blinding to the other assessor's grading ensured by undertaking the grading when the other rater was not present.

The raw demographics data of the sub-set of patients is shown in appendix 4. The sample comprised 13 male and 17 female subjects whose symptoms were on the left side in 12 cases and right side in 18 cases. The type of trial were counterbalanced to give 15 loaded and 15 non-loaded trials. To allow for a degree of comparability with the inter-rater reliability sample examined by McClure et al. (2009), BMI and maximum range of movement on the symptomatic side from the goniometric data for each subject is presented below.

Table 4-6: Demographics of scapular dyskinesis test intra-rater reliability study subjects

Variable	Measure of variability			
	Mean	SD	Minimum	Maximum
BMI (kgm^{-2})	29.0	6.3	18.2	50.5
Maximum range of movement on symptomatic side (°)	138	25	74	174

Key: Kg = kilograms, m = metres, BMI = body mass index, ° = degrees, SD = standard deviation

It can be seen from the above table that the average BMI was 29.0 kgm^{-2} and from the raw data that 13 of the 30 subjects (43%) had a BMI of greater than or equal to 30.0 kgm^{-2}. There was wide variation in the maximum range of movement on the symptomatic side around a mean of 138.0°.

The raw data of the scapular dyskinesis gradings for each rater is in appendix 4 and the frequency of the gradings shown in the below table.

Table 4-7: Frequency of the Scapular gradings from the inter-rater reliability study (n=30)

	Scapular Dyskinesis Grading		
Rater	Normal	Subtle abnormality	Obvious abnormality
MS	8	13	9
KN	5	10	15

Table 4-8: Cross-tabulation table of Scapular gradings from the inter-rater reliability study (n=30)

		Rater is KN		
		Normal	Subtle abnormality	Obvious abnormality
Rater is MS	Normal	3	1	4
	Subtle abnormality	1	8	4
	Obvious abnormality	1	1	7

Assessment of agreement between MS and KN was performed using Cohen's Kappa (McHugh 2012)

Table 4-9: Inter-rater reliability of Scapular gradings

n=	Kappa	p value
30	0.395	0.002

For the Scapular gradings the Kappa value showed fair agreement (Kappa 0.21 to 0.40) (Viera and Garrett 2005), borderline moderate (Kappa 0.41 to 0.60) (Sim and Wright 2005) between MS and KN. Based upon the Kappa value it can therefore be concluded that when compared to an experienced musculoskeletal outpatient physiotherapist involved with the study, MS demonstrated a fair level of agreement for scapular dyskinesis test gradings.

To allow for more direct comparability with the original publication by McClure et al. (2009), a linear weighted Kappa was also performed. Although McClure et al. (2009) didn't state the scaling of their weightings, it was assumed that it was 0, 1, 2 for same grade, one grade disparity and two grades disparity; respectively.

Table 4-10: Inter-rater reliability of Scapular gradings; linear weighted Kappa

n=	Kappa	p value
30	0.329	0.002

4.1.3.1.1 Interpretation

The linear weighted Kappa results confirm fair agreement (Kappa 0.21 to 0.40) (Viera and Garrett 2005) for inter-rater reliability of Scapular grading. Therefore the SDT must be applied with caution in the main cohort study.

4.2 E Structural pathology

4.2.1 E1. Structural pathology via imaging

In this prognostic cohort study, structural pathology was assessed as a potential prognostic variable. Specifically, diagnostic ultrasound was used to image the rotator cuff tendons and subacromial bursa. Full details of how the imaging was performed and the diagnoses determined are provided in the method section (page 131). However as a highly operator dependent modality with variable levels of agreement in the published literature (Ottenheijm et al. 2010) it was deemed necessary for the MS to undergo formal training in diagnostic ultrasound and to undertake an inter-rater reliability study.

4.2.1.1 Clinical training in diagnostic ultrasound

MS is a physiotherapist by clinical training and prior to commencement of the study had no diagnostic imaging experience. Through the School of Medicine in Cardiff University a 2 year Postgraduate Certificate (PgC) in Medical ultrasound was undertaken as part of the training component of his PhD. The training was extensive and included formal assessments of theoretical knowledge, image optimisation and sonographic diagnosis (http://www.case-uk.org/; accessed 24/04/2017). Further details of the course requirements are provided in appendix 4. MS successfully completed the PgC (Distinction grade) and is on the voluntary register of Sonographers operated by the Society of Radiography (http://www.sor.org/practice/ultrasound/register-sonographers ; accessed 24/04/2017).

4.2.1.2 Shoulder ultrasound reproducibility study

For the purposes of the PhD it was essential that the ability of MS to accurately diagnose structural pathology was assessed. To do this, comparison with directly observed pathology on arthroscopy would have provided a gold standard comparator (Jeyam et al. 2008). However this would lack transferability to the target sample because the prognostic cohort study sample was drawn from a non-surgical population. Alternatively, comparison could have been undertaken against another imaging modality such as MRI (Vahlensieck 2000) but for cost reasons the inclusion of additional MRI scanning would have been prohibitive.

A reproducibility study was therefore undertaken comparing scans performed by MS against an experienced musculoskeletal ultrasound practitioner, Dr Peter Mullaney (PM). PM has been a Consultant Radiologist since 2007 and is a level 3 ultrasound practitioner (The Royal College of Radiologists 2012). Scanning was undertaken on a convenience sample of patients referred for ultrasound scans of their shoulder in the Radiology department at the University Hospital of Wales (UHW). For pragmatic reasons no restriction was placed upon the pathology the patient was referred with. Radiology department scanning was performed on a Toshiba Aplio 500 unit (Toshiba Medical Europe B.V.) using high frequency linear musculoskeletal probes, depending on body habitus.

Potential subjects were sent a patient information sheet (appendix 4) along with their Radiology appointment letter. The option to include a particular subject in the study was dependent upon clinical service demands during that particular clinical session. Where service demands allowed, MS approached potential subjects in the waiting area to answer any questions they had and as appropriate took consent using the consent form in appendix 4.

4.2.1.3 Performance of ultrasound scan for reproducibility study

All shoulder ultrasound scans were based upon the internationally recognised protocol published by the European Society of Skeletal Radiology (Beggs et al. 2010; appendix 4) to ensure that each scan was performed in a standardised manner. This included all structures around the shoulder and reflects routine sonographic practice for the shoulder (Corazza et al. 2015). However for the purposes of this study, only those structures directly implicated in the theorised pathological models of SIS/RCTendinopathy were compared for sonographic reproducibility. These were cuff tears and subacromial bursal involvement along with tendinopathic findings and calcific deposits (Read and Perko 1998).

Each patient was first scanned by MS who wrote down the sonographic findings, specifically in relation to the components of the rotator cuff (Subscapularis, Supraspinatus, Infraspinatus/Teres Minor) and the subacromial bursae (Smith et al. 2015). For the sonographic diagnostic criteria, these were based upon the clinical practice of PM and the work of McNally (2014). The same patient was then immediately scanned by PM who was the senior clinician with overall responsibility for the sonographic diagnosis. Differential sonographic diagnoses and proforma for data collection were based upon section 1b and section 1c of the ultrasound differential diagnoses and record scan findings in appendix 4 (Smith et al. 2015).

4.2.1.4 Analysis and interpretation of ultrasound reproducibility study data

Findings for each component of the rotator cuff (Subscapularis, Supraspinatus, Infraspinatus/Teres Minor) were collapsed to provide a single sonographic diagnosis judgement in relation to the presence of each of tendinopathic change, calcific deposits, bursitis, partial and full thickness tears. Where one or more components of pathology was identified in the cuff (independent of bursal findings) then a 'cuff pathology' diagnosis was assigned.

4.2.1.4.1 Demographics of ultrasound inter-rater reliability study subjects

The ultrasound inter-rater reliability study was undertaken between 21/8/2013 and 21/5/2014 and the demographic and clinical data can be seen in appendix 4. The sample comprised 15 male and 20 female subjects with symptoms on the left side (n=12) compared to the right (n=23). Referral source was predominantly via GP (n=30) with two from Rheumatology, two from trauma / fracture clinic and one not recorded. The patient reported reason for onset of symptoms was insidious (n=22), trauma (n=5), combination of trauma and insidious (n=4), Rheumatoid Arthritis (n=2), overuse (n=1) and not recorded (n=1). Data on specific shoulder pathology was not recorded.

Table 4-11: Demographics of US inter-rater reliability study subjects

Variable	Measure of variability			
	Mean	SD	Minimum	Maximum
Age (years)	54.0	15.9	21.0	82.0
Chronicity (months)	23.9	38.2	1.0	216.0

Key: SD = standard deviation

The raw data and measures of variability demonstrate that there was wide variation in the age and chronicity of the subjects in the ultrasound inter-rater reliability study. This mirrors the non-specific nature of the shoulder patients seen in the Radiology department, which were predominantly from GP referrals, with a predominance of insidious onset.

4.2.1.4.2 Inter-rater reliability of sonographic diagnosis

The raw structural pathology data from the ultrasound inter-rater reliability study are shown in appendix 4 and the incidence of each diagnostic category per rater is shown in table 1-12:

Table 4-12: Incidence of each diagnostic category per rater

Sonographic diagnosis	Rater = MS		Rater = PM	
	Finding recorded? Yes n=	Incidence of pathology identified n= ; (% of Finding recorded)	Finding recorded? Yes n=	Incidence of pathology identified n= ; (% of Finding recorded)
Cuff pathology	34	23 (68%)	35	22 (63%)
Tendinopathy	34	20 (59%)	35	21 (60%)
Calcific deposits	33	5 (15%)	35	2 (6%)
Bursitis	33	19 (58%)	19	6 (32%)
PTT	34	1 (3%)	35	3 (9%)
FTT	34	4 (12%)	35	3 (9%)

Key: PTT = Partial Thickness Tear; FTT = Full Thickness Tear

As per the raw data in appendix 4, one or more pairs of data in each sonographic diagnosis category was missing which is why although the number of inter-rater reliability subjects was 35, the number of recorded sonographic diagnosis pairs was 34 or less in each category.

The data in the above table demonstrate that tendinopathy and bursitis were the most frequently assigned individual pathologies by each rater whilst the composite grading of 'cuff pathology' had the highest overall incidence.

Assessment of agreement between MS and PM was performed using Cohen's Kappa (McHugh 2012) where the diagnosis provided by PM was defined as the gold standard. In addition sensitivity and specificity values were calculated where sensitivity = true positive / (true positive + false negative) and specificity = true negative / (true negative + false positive) (Lalkhen and McCluskey 2008).

Table 4-13: Inter-rater reliability of sonographic diagnosis

Sonographic diagnosis	n=	Kappa	Kappa 95% C.I.	p value	Sensitivity (%)	Specificity (%)
Cuff pathology	34	0.358	0.032 to 0.647	0.035	53.8	81.0
Tendinopathy	34	0.150	-0.174 to 0.481	0.382	65.0	50.0
Calcific deposits	33	0.298	0.000 to 0.784	0.016	100.0	87.5
Bursitis	19	0.503	0.191 to 0.791	0.012	100.0	61.5
PTT	34	0.477	0.000 to 1.000	0.001	33.3	100.0
FTT	34	0.841	0.000 to 1.000	0.000	100.0	96.8

Key: PTT = Partial Thickness Tear; FTT = Full Thickness Tear

The data in the above table demonstrate a non-significant Kappa value for tendinopathy. For the sonographic diagnoses of cuff pathology and calcific deposits the Kappa values showed fair agreement (Kappa 0.21 to 0.40) (Viera and Garrett 2005) with high specificity for both cuff pathology and calcific deposits. Maximum sensitivity was demonstrated for calcific deposits.

For bursitis, agreement was in the moderate range (Kappa 0.41 to 0.60) (Viera and Garrett 2005), although as almost half of the data set pairs had no bursal findings recorded by one of the raters, caution must be applied when interpreting the Kappa value (Adejumo 2005). Maximum sensitivity was demonstrated but lower specificity.

For PTT and FTT an almost complete data set was available with Kappa in the moderate range (Kappa 0.41 to 0.60) and almost perfect agreement range (0.81 to 0.99) (Viera and Garrett 2005); respectively. For PTT the low sensitivity indicates that MS's ability to identify PTT tears that were present was low. However the ability to correctly exclude pathology was absolute. With FTT MS's ability to identify PTT tears that were present was perfect and the ability to correctly exclude pathology was very high. However the low incidence of PTT (n=3) and FTT (n=3) in the Inter-rater reliability sample means that such values must be viewed with caution.

Based upon the Kappa values it can therefore be concluded that when compared to an experienced musculoskeletal radiologist, MS demonstrated acceptable agreement for all categories except for the sonographic diagnosis of tendinopathy.

4.3 Treatment

4.3.1 Development of treatment recording tool

One potential approach to recording the nature of the clinician-directed intervention received by patients was for MS to transcribe from the treatment notes onto a standardised form. However, this risked potential incorrect interpretation of clinician notes and subsequent errors with recording the treatment.

Instead, a bespoke treatment-recording tool was developed in collaboration with the treating clinicians. The preliminary version in appendix 4 was developed by MS following identification of common categories of treatment from the literature. However feedback from the treating clinicians identified the following problems:

- Some of the gross categories had more emphasis, e.g. three entries for exercise therapy, two for stretching and one for the other categories. This carried a potential risk for biasing clinicians in the modalities they delivered.
- Recording the time spent on the modality during treatment session was considered to be too time consuming for clinicians to complete and prone to error in estimating.
- The option for all modalities to be given as a home exercise programme was of limited relevance for some options, e.g. manual therapy, electrotherapy and taping.

Following feedback the final version in appendix 4 was developed. The cover page provided a brief context for the study and emphasised the non-judgmental, anonymous nature of the treatment data collection and analysis.

Key elements of the final version were:

- Use of circling pre-completed boxes so as to make it more user friendly and quicker.
- Similar categories used as in the preliminary version but with an equal weighting for each type.
- Prompts given to encompass the main sub-elements and use of highlighting to differentiate the different exercise therapy elements.
- Separate section for home exercise programme.
- An 'Additional details' section to capture any modalities not included on the form.

All methods by which the cohort study was undertaken will be detailed in the following Methods chapter.

5 METHODS

This chapter presents the methods by which the prospective cohort study was undertaken. This will comprise the study design, ethical approval and aspects of the pathology, patient care pathway, clinical setting and sample size. The procedures by which the patient sample was identified and recruited will be presented followed by each of the potential prognostic factors and how they were measured. The method by which the pragmatic treatment was captured will be explained along with the defining and measurement of the outcome. The prognostic variable trimming approach will then be presented along with the prognostic data analysis approach. Aligning with the STROBE statement (http://www.strobe-statement.org. ; accessed 25th January 2018) regarding the reporting of observational studies, these map to the sections of this chapter as per the below table.

Table 5-1: Alignment of methods chapter with components of STROBE

Component of STROBE	Section in methods chapter
Study design	5.1 Study design
Setting	5.3.3 Patient care pathway and clinical setting
Participants	5.3.1 Clinical presentation of SIS/RCTendinopathy 5.4 Identification and recruitment of patient sample
Variables	5.5 Potential prognostic factors and how they were measured 5.6 Treatment / intervention 5.7 How outcome is defined and measured
Data sources/ measurement	5.5 Potential prognostic factors and how they were measured 5.6 Treatment / intervention 5.7 How outcome is defined and measured
Bias	5.7 How outcome is defined and measured
Study size	5.3.4 Sample size
Quantitative variables	5.8 Prognostic variable trimming approach
Statistical methods	5.9 Prognostic data analysis approach

5.1 Study design

This was a prospective cohort study investigating potential prognostic factors in patients with SIS/RCTendinopathy undergoing physiotherapy with outcome defined by change in a functional outcome score (OSS).

The potential prognostic factors will be selected across a range of categories as presented in the literature review (A. to E2.). Treatment was pragmatic but was not included within the prognostic model as only baseline potential prognostic factors were considered. Blinding of MS to the OSS was

ensured by placing each outcome score in a sealed envelope and retaining it thus until all other variables were processed and analysed.

5.2 Ethical approval

Ethical approval for the study was via an amendment to a pre-existing umbrella ethical approval (REC reference: 10/MRE09/28) and NHS R&D approval was secured under C&V UHB reference: 12/OAE54/48.

5.3 Pathology, patient care pathway and clinical setting

Subjects were required to have a clinical diagnosis of SIS/RCTendinopathy and been referred for outpatient physiotherapy in C&V UHB. No restrictions were placed upon the clinical source of the referral.

5.3.1 Clinical presentation of SIS/RCTendinopathy

As noted previously the commonly used clinical tests for SIS/RCTendinopathy have poor sensitivity or specificity (Calis et al. 2000). Furthermore, there is controversy as to the signs and symptoms of SIS/RCTendinopathy (Lewis 2009). Due to the poorly differentiated nature of the pathology, patients were identified via a process of elimination of pathologies discrete from SIS/RCTendinopathy. This was supplemented by the treating clinician identifying any subject who was subsequently found to have a non-SIS/RCTendinopathy pathology.

5.3.2 Inclusion and exclusion criteria

Table 5-2: Inclusion and exclusion criteria

Inclusion criteria for SIS/RCTendinopathy sample	Exclusion criteria for SIS/RCTendinopathy sample
• 18 years of age or older • Referred for out-patient physiotherapy in C&V UHB • Ability to provide informed, written consent • Clinical presentation of local shoulder pain:	• Shoulder surgery on the involved side in the previous year • Trauma leading to a sudden onset of symptoms within the last 4 weeks • Referred following a fracture of the shoulder girdle or upper limb

○ Pain in the GHJ region or proximal segments of the arm ○ Reproduction or worsening of symptoms upon use of the arm.	• Reproduction of shoulder symptoms from the cervical spine • Rheumatoid arthritis or similar inflammatory joint disorder • Imaging confirmed massive rotator cuff tear • Clinical presentation of shoulder pathology discrete from SIS/RCTendinopathy; specifically: ○ Shoulder instability ○ Adhesive capsulitis ○ ACJ pain ○ Radiologically confirmed osteo-arthritis • Multi-area symptoms

Patients were excluded if they were referred following shoulder surgery because the study was concerned with identifying the characteristics of those patients who are treated conservatively (i.e. with physiotherapy) as oppose to a composite of surgery and post-operative non-invasive, multimodal physiotherapy.

Trauma leading to a sudden onset of symptoms was an exclusion criterion, as this was considered potentially indicative of a massive rotator cuff tear (Turman et al. 2010). However an exclusion timescale was capped at it occurring within the last 4 weeks. This was applied so that where trauma was an incidental or contributing factor to the development of overuse / impingement related symptoms then such patients could still be included in the study.

Patients were excluded if they were referred following a fracture of the shoulder girdle or upper limb because the study was concerned with the soft-tissue pathology of SIS/RCTendinopathy as oppose to bony fractures. Whilst it is acknowledged that SIS/RCTendinopathy could still be present in a patient with a recent fracture there are differences in clinical management and treatment timescales for post-fracture patients (Rangan et al. 2015).

Where the cervical spine was identified as the source of the symptoms experienced in the shoulder then such patients were excluded as this was considered to be discrete from SIS/RCTendinopathy

(Gorski and Schwartz 2003). However, patients with concomitant cervical spine symptoms were not excluded. Similarly, where a systemic pathology such as rheumatoid arthritis was identified as the source of the symptoms experienced in the shoulder then such patients were excluded as this was considered to be discrete from SIS/RCTendinopathy (Smith et al. 2005). Imaging confirmed massive rotator cuff tears were excluded as conceptually this was deemed a substantial progression from the target population of SIS/RCTendinopathy with a discrete conservative management approach (Ainsworth 2009).

Patients were excluded if they were referred following shoulder subluxation or dislocation because the study was concerned with the overuse / impingement related pathology of SIS/RCTendinopathy. Evident disruption of the GHJ was considered a separate clinical entity with clear differences in clinical management and treatment timescales (Hayes et al. 2002).

Adhesive capsulitis was defined as being a substantial restriction in shoulder range of movement (particularly external rotation) with or without the presence of shoulder pain (Struyf and Meeus 2014). Although there is uncertainty as to its aetiology and a degree of overlap is noted with SIS/RCTendinopathy, it can be regarded as being a separate clinical entity from SIS/RCTendinopathy (Zuckerman and Rokito 2011).

An isolated ACJ problem was defined as instability of the ACJ following trauma and/or pain locally reproduced at the ACJ (Armstrong 2014). Severe osteoarthritis confirmed radiographically typically presents from the 6th decade onwards with associated restricted ROM and pain (Armstrong 2014).

Patients were excluded if the referral included multiple anatomical regions of musculoskeletal symptoms because for service delivery reasons the shoulder-specific treatment for such patients would inevitably need to be balanced with the treatment for the other anatomical regions. This could lead to skewing of the treatment duration and total number of appointments data for such patients.

Where baseline function (as measured by the primary outcome measure of OSS) was within the minimally clinically important change (MCIC) of 5 (Ekeberg et al. 2010b) of maximum then such patients were excluded from the final analysis because conceptually no ‘improvement’ could occur. However as MS was blinded to the OSS scores, this exclusion criteria could not be applied until all data collection was completed.

5.3.3 Patient care pathway and clinical setting

The inclusion criterion of patients being referred for out-patient physiotherapy in C&V UHB meant that they were recruited at the same place within the SIS/RCTendinopathy care pathway, namely upon referral to secondary care based physiotherapy. However as no restriction was placed upon who or how patients were referred in, the specific component of the patient management pathway this represented (i.e. following consultation and/or treatment by General Practitioner, Orthopaedics, Rheumatologist or other speciality) was not pre-defined. However all patients were recruited and baseline data collected immediately prior to their first clinical contact. As such, the cohort can be considered in a limited capacity to be an inception cohort.

Data collection occurred at two different sites (University Hospital Wales (UHW) and Whitchurch hospital (WHI)) which were a major, regional centre and a district general hospital type (DGH) setting; respectively. This was to address the potential limitation of data collection only occurring at a specific type of setting, thereby limiting transferability of the findings.

5.3.4 Sample size

As noted in the literature review, sample size calculation procedures are an area of controversy within the prognostic literature (Riley et al. 2013). For the purposes of this exploratory prognostic study, the guide of 10 cases per variable was employed (Peduzzi et al. 1996). As per the literature review structure of potential prognostic factors, 10 variables were identified (across the five categories), giving a sample size requirement of 10 x 10 = 100.

5.4 Identification and recruitment of patient sample

The mechanism by which the patient sample was identified and recruited is described below.

5.4.1 Stage 1: geographical location

Potential subjects were required to have been referred for outpatient physiotherapy within C&V UHB. This is a large out-patient physiotherapy service covering the capital city of Wales, with 2010 figures showing that over 23,000 musculoskeletal patients were treated across the 7 C&V UHB out-patient physiotherapy sites (Cardiff & Vale University Health Board 2015a).

In C&V UHB, referrals are received centrally and assigned to specific clinical sites according to the geographical area where the patient lives. Two outpatient physiotherapy sites were selected for data

collection, thereby forming the first stage of identifying the patient sample as the geographical area where a potential subject lived would determine whether or not they would be screened for inclusion in the study.

5.4.2 Stage 2: anatomical location of symptoms

Identification of potential subjects with shoulder problems was performed by MS searching the electronic waiting list database. This database was screened on a weekly basis by site (UHW, WHI) and anatomical area (Shoulder) and stratified according to time on the waiting list. This allowed MS to recruit from the bottom of the waiting list, thereby minimising the delay between initial contact with potential subjects and subsequent data collection.

5.4.3 Stage 3: screening of written referral for SIS/RCTendinopathy

For each shoulder patient who was identified as a potential subject, MS located the scanned electronic copy of their referral for outpatient physiotherapy. MS is an experienced musculoskeletal physiotherapist with 15 years' experience and screened each written referral using the previously presented inclusion and exclusion criteria.

5.4.4 Invitation of potential subjects to participate in the study

Once a shoulder pain patient had been identified as a potential subject, MS sent them a study invitation pack on behalf of the strategic lead for physiotherapy. The contents of this postal invitation are shown in appendix 5 and comprised:

- Letter outlining the study, inviting the potential subject to participate and directing the potential subject to the self-screening form
- Self-screening form
- Sealed envelope containing:
    - Patient information sheets
    - Set of baseline questionnaires

5.4.5 Stage 4: decision by potential subject of whether to participate in study and subsequent self-screening

The self-screening form in appendix 5 was adapted with kind permission from one developed by Dr Rachel Chester. At the time of the data collection Dr Chester was also undertaking a study of prognostic indicators in patients with shoulder pain receiving physiotherapy.

The self-screening form enabled those patients who were considering participating in the study to self-identify if they were grossly eligible for the study. This involved using lay terminology to convey the inclusion and exclusion criteria of the study. Specifically “Is stiffness in your shoulder much more of a problem for you than pain in your shoulder?” was used in an attempt to exclude adhesive capsulitis by emphasising the stiffness component of the pathology. “Do your shoulder symptoms get worse when you move your neck rather than your shoulder in a particular position?” was used in an attempt to exclude the cervical spine as the source of the symptoms experienced in the shoulder. However it is noted that the effectiveness of the self-screening process was dependent upon the potential subjects (who were often lay individuals) correctly interpreting the descriptors.

The above processes meant that potential subjects would not be recruited if they decided that they did not want to be involved in the study or self-identified that they were in-eligible for the study. As the postal information did not direct such patients to make further contact with MS then it was not possible to differentiate the reason for that potential subject not being recruited. Consequently the rate of incorrect self-exclusion could not be determined.

5.4.6 Stage 5: telephone screening of potential subject

Those potential subjects who were happy to participate in the study and self-identified as being eligible contacted MS by phone or email. This enabled MS to use his clinical experience to repeat the screening process with a refined level of triage. Where a potential subject was identified as being eligible, MS provided additional information on the study and answered any questions that the subject had. MS also reminded the subject to bring the questionnaires along when they attended and to complete them as close to the time of the appointment as possible. This minimised the potential for these inherently subjective potential prognostic factors to be biased by clinician interaction and also made for an efficient use of the subject’s time on the day of their first appointment.

To ensure accurate but anonymous tracking of patients throughout the study, each was assigned a patient specific code which comprised their gender (M for male, F for female), the last 4 digits of their hospital number (this was prefixed by TM if this was a temporary hospital number) and then h (for UHW) and w for WHI.

5.4.7 Logistical challenges associated with recruitment and attrition

As per the letter outlining the study and inviting the potential subject to participate, potential subjects who had decided they were happy to participate in the study and had self-identified as being eligible were directed to contact the physiotherapy Department to make an appointment.

Baseline data collection had to occur immediately prior to the patient's first appointment so that (i) capture of the potential prognostic factors truly reflected the status of the patient at the commencement of physiotherapy and (ii) the potential prognostic factors were not biased by patient-clinician interaction, including the assessment process. However the study had to assimilate into routine clinical care and this was ultimately determined by clinical service need, which meant that clinical service provision procedures and pressures were the ultimate determinants of if – and when – data collection could occur.

In order to maximise efficient recruitment to the study MS took a multi-layered approach to interacting with both potential subjects and the receptionist staff:
(i) potential subjects were contacted via text or telephone call as a follow up to the initial invitation letter. Up to two texts or voice messages were used before attempting to recruit a potential subject was ceased

and

(ii) weekly screening of the electronic clinical records for all potential subjects was performed to identify if they had been removed from the waiting list due to no contact OR had already been assigned a clinical appointment

and

(iii) weekly screening of the physiotherapy clinical appointments for each site to identify if a potential subject had been booked in for data collection but MS had not been informed.

The time period over which recruitment occurred and the recruitment numbers will be presented in the results section.

5.5 Potential prognostic factors and how they were measured

The mechanisms by which the potential prognostic factors from categories A. to E2. were collected will now be presented. For variables A. demographics and B. clinical history, these were collected using the patient data collection form shown in appendix 5.

5.5.1 A. Demographics

As noted in the literature review, a wide variety of demographic related variables have been considered in previous prognostic studies. For the purposes of this study, data across these variables was collected in order to adequately describe the sample. However, in the data reduction section of the results chapter the choice of variable to be entered into the prognostic model will be presented.

Generic demographic variables were:

- Age at time of first appointment, calculated from date of birth. Retained as a continuous variable
- Gender; taken from clinical records
- Smoking history; patient reported as (i) current, (ii) previous, (iii) never
- Education level; identified as (i) College / University, (ii) less than 12 years in education (Engebretsen et al. 2010)
- BMI calculated from height and weight recorded at first appointment

Work, recreational or functional task related variables:

- Currently working; patient reported as (i) full time, (ii) part time, (iii) retired, (iv) full time carer, (v) part time carer, (vi) student or unemployed
- Paid work type; patient reported as (i) not paid working, (ii) professional / managerial / office, (iii) manual / semi-manual / unskilled
- How often carry 10Kg at work; patient reported as (i) seldom / never, (ii) sometimes, (iii) extremely / often, (iv) not applicable as not working (Reilingh et al. 2008)
- How often work above shoulder height; patient reported as (i) seldom / never, (ii) sometimes, (iii) extremely / often, (iv) not applicable as not working (Reilingh et al. 2008)
- Play overhead sports; patient reported as Yes / No

5.5.2 B. Clinical history

As with the demographic variables, a wide variety of clinical history related variables have been considered in previous prognostic studies and these partially informed the collection of such variables in the current study.

For the purposes of this study, data across these variables was collected in order to adequately describe the sample. However, in the data reduction section of the results chapter the choice of variable to be entered into the prognostic model will be presented.

Current condition related variables:

- Source of referral
- Duration current episode; patient reported as (i) 0 to <3 months; (ii) 3 to <6 months; (iii) 6 to <12 months; (iv) 12 to <24 months; (v) 24 months or longer (Engebretsen et al. 2010)
- Involvement of dominant side; patient reported as Yes / No
- Precipitating cause; patient reported as (i) Unknown / insidious / gradual onset; (ii) Injury / trauma; (iii) Strain / overuse: unusual activities; (iv) Strain / overuse: usual activities
- Previous shoulder pain; patient reported as Yes / No
- Associated neck pain; patient reported as Yes / No
- Neurological symptoms; patient reported as (i) None; (ii) Pins and needles / tingling; (iii) Distal symptoms; (iv) other
- Course of condition; patient reported as (i) Worse; (ii) Same; (iii) Better

Treatment variables were:

- Previous rehabilitation; patient reported as (i) None; (ii) Physiotherapy; (iii) Chiropractic; (iv) Massage
- This episode; currently taking analgesia/NSAIDS, patient reported Yes / No
- This episode; treatment via GP; patient reported as (i) No contact; 2 = Advice only; 3 = oral or topical medication (Analgesia / NSAIDs) (+/- advice); 4 = oral medication + injection; 5 = injection only; 6 = referral only
- This episode; patient reported number of GP injections
- This episode; injection by non GP, patient reported Yes / No
- This episode; investigations, patient reported (i) None; (ii) U/S; (iii) MRI; (iv) X-ray; (v) U/S and X-ray

5.5.3 C1. Patient reported measures: Pain

As per appendix 5 ‘Order of testing’, patients completed the VAS shown in appendix 5 ‘Visual Analogue Scale’ prior to undertaking any physical assessment or testing so as to minimise any contamination of this variable by the physical procedures. Pain was assessed at baseline for all prognostic cohort study subjects using the 100 mm horizontal line VAS. This was anchored with the phrases ‘no pain’ and ‘worst pain imaginable’ (Jensen et al. 1986; Hawker et al. 2011). Written instructions as to the context of the measure and how to complete it were provided on the form so

as to standardise its completion; MS was present to answer any patient queries. Pain under three conditions was recorded: 'On activity', 'At rest' and 'At night over the last week'.

'On activity' was selected in order to capture the dynamic, functional and movement related elements of SIS/RCTendinopathy (Schellingerhout et al. 2008). Patients were directed to consider a movement or activity that reproduced their pain as the reference for this pain rating. 'At rest' was selected in order to capture the pain levels when the arm was not being used, the prompt being "for example when you are just resting with your arm by your side". As with 'On activity', for the 'At rest' score, patients were directed to rate this as a momentary rating rather than one over a period of time.

'At night over the last week' was selected in order to capture the commonly reported sleep-related and night time pain (Factor and Dale 2014). As this was likely to incorporate numerous elements such as prolonged periods of non-movement, different sleeping postures and turning in bed then patients were directed to rate this over a period of time, i.e. 1 week.

Data processing: VAS readings were recorded for each condition, namely 'On activity', 'At rest' and 'At night over the last week'. A mean score was calculated to provide a fourth pain variable. The linear measurement was undertaken by MS and re-checked by a colleague (Steve Hiles; SH) for quality control purposes. Where one or more of the individual scores was missing then a mean score was not calculated.

5.5.4 C2. Patient reported measures: Psychological symptoms

As per appendix 5 'Order of testing', the psychological symptoms questionnaire (4DSQ) was posted to patients for completion immediately prior to their first appointment.

The 4DSQ (available from www.emgo.nl/researchtools/4DSQ.asp ; appendix 5 '4DSQ') is a self-contained questionnaire with instructions as to the background to the questionnaire, the timescale over which responses should be derived and how to complete it. In addition, MS was available at the first appointment to answer any queries that patients had about completing the questionnaire.

At the time of data collection, MS reviewed the scores and where elevated readings were present MS encouraged the patient to contact their GP to discuss their psychological well-being. This was

separate to the purposes of the research study and was undertaken from the perspective of duty of care to the patient.

Data processing: the 4DSQ scores were calculated using the procedures detailed in appendix 5 '4DSQ: Scoring and interpretation' whereby each question relates to a particular subsection (Distress, Depression, Anxiety or Somatisation). The responses are scored according to the frequency with which the patient reports experiencing those symptoms. Collapsing down the 'regularly', 'often' and 'very often or constantly' to the same score means that emphasis is placed on the number of symptoms rather than their severity or constancy *per se* (Terluin et al. 2006). The scores for each subsection are then totalled and cut off points used to identify moderately elevated or strongly elevated symptoms for that subsection.

For the purposes of this study the originally published cut off points (Terluin et al. 2006) were not applied. Instead the adjusted anxiety cut off points subsequently published and advocated by the original developers was used. This was based on data from Terluin et al. (2014a) and the higher cut-off point for anxiety was further adjusted as per (Terluin et al. 2014b) to align with its use in English.

5.5.5 C3. Patient reported measures: Function / Disability

The Function / Disability questionnaire (SPADI) was posted to patients for completion immediately prior to their first appointment. The wording and guidance provided to respondents was standardised as per appendix 5 'SPADI', whereby the second page was sent to patients to complete and the first page used by MS for data processing.

Data processing: each of the 13 questions is scored out of 10 with 0 being the least and 10 being the worst pain or difficulty. The five pain questions give a pain score via the score in points divided by 50 and multiplied by 100 to give a percentage score. The eight disability questions give a disability score via the score in points divided by 80 and multiplied by 100 to give a percentage score. The total SPADI score is derived by summing all the scores, dividing by 130 and multiplying by 100 to give a percentage score. Guidance is also provided for any missing values although MS was available to answer any queries that patients had about completing the questionnaire.

5.5.6 D1. Clinical measures: Strength

Isometric strength was assessed at baseline for all prognostic cohort study subjects. As with previous prognostic studies that used a hand-held dynamometer (Hung et al. 2010; Chester et al. 2016),

internal rotation and external rotation were measured. Whilst such studies also measured abduction, elevation in the scapular plane was performed in the current study (Hsu et al. 2009).

5.5.6.1 Patient positioning and testing

All testing was performed with subjects seated as this prevented them from using their body weight to generate additional force; it also ensured that subjects were stable during testing. IR and ER testing were performed with the arm against the side (i.e. GHJ in 0° flexion and 0° abduction). Neutral GHJ rotation was ensured by having the elbow flexed at 90° and the fingers pointing forwards (Cools et al. 2014). Forearm pro/supination was standardised by having the palm facing upwards. Scaption testing position was standardised by flexing the GHJ to 90° in the scapular plane and by having subjects lead with their thumb uppermost (Harrington et al. 2013). For reasons of time pressure, testing positions were confirmed visually rather than goniometrically.

5.5.6.2 Testing procedure

Due to the unfamiliar nature of the isometric testing procedure it was first verbally explained to subjects while MS demonstrated the test procedure on himself. For IR and ER the subjects performed the physiological movement to familiarise themselves with it. For IR, ER and scaption the isometric contraction was then performed with the static nature emphasised but without resistance being applied. For IR the verbal cue “keep your elbow tucked into the side and try to move your hand across your body” and for ER “keep your elbow tucked into the side and try to move your hand away from your body” were used so as to minimise the potential for subjects to attempt to perform isometric adduction or abduction; respectively.

Demonstration was performed on the asymptomatic side (where bilateral symptoms were present the least symptomatic side was used) and then the symptomatic side so that symptom exacerbation due to performance of an unfamiliar test was minimised. Testing was performed in the same order of IR then ER then scaption so as to limit pain inhibition affecting the strength readings. This order was chosen because clinical experience indicated that symptom exacerbation was likely to be greatest with scaption as this closely resembled one of the orthopaedic symptom reproduction test positions (Michener et al. 2009; Hegedus et al. 2012). The order of side testing (right then left) was the same throughout.

The verbal instructions given to subjects was to push as hard as they could within the limits of their pain. Subjects self-determined how hard to push and no attempt was made to quantify the ‘within

limits of pain' instructions (e.g. via a visual analogue score rating) due to the potential for confusion if subjects were to be asked to perform the isometric contraction and concurrently rate their pain.

MS took the measures in a stride-standing position with his hand holding the dynamometer in contact with his torso. This enabled him to offer maximal, stable resistance to the contraction of the subject (Lu et al. 2007). Resistance was applied at the distal ulnar (IR) and distal radius (ER and scaption) (see page 413) and three measures recorded (i.e. three trials) for each side at each position. The unit recorded the maximal force (Newtons) generated for each trial. The distance from elbow to distal radius (IR and ER) and GHJ to distal radius (scaption) was recorded using a tape measure so that the force (Newtons) could be converted to a rotational moment (Newton metres).

5.5.6.3 Data processing

The mean force (Newtons) of the three trials per side at each position was calculated and converted to a rotational moment (Newton metres) by multiplying the mean force by the moment arm (metres) (Schrama et al. 2014). Even though the protocol included opportunity for the subjects to familiarise themselves with the testing procedure, there was still the potential for learning to occur and/or symptom exacerbation, either of which had the potential to affect an individual trial. To minimise the impact of these factors, the mean of the three trials was selected as the variable of interest.

The strength measures in the prognostic cohort study are not presented relative to the dominant side but instead relative to the symptomatic side as this is of greater clinical relevance. It is acknowledged that the strength intra-rater reliability study data however was presented relative to the dominant side. This was deemed a relevant method of analysis because the intra-rater reliability study subjects were asymptomatic, i.e. they did not have a symptomatic / asymptomatic side. Furthermore, the intra-rater reliability study data (ICC(3,1)) was presented per side as this allowed for confirmation that single side measures were robustly recorded.

5.5.7 D2. Clinical measures: ROM

Active, planar ROM (elevation in the scapular plane, i.e. Scaption) was assessed at baseline for all prognostic cohort study subjects.

5.5.7.1 Patient positioning and testing

Scaption ROM was measured with subjects in standing with their knees and hips in a comfortable, stable position. Scaption was defined as 30° anterior to the coronal plane (McClure et al. 2006). In order to standardise the plane of elevation between subjects, vertical poles were used and a template with a 30 degree angle marked on it was used to ensure the vertical poles were at a 30 degree angle relative to the foot positions. Subjects were instructed to stand on the foot markers and elevate their arms bilaterally, keeping their elbows fully extended, thumbs pointed upwards and their arms lightly in contact with the vertical bars (McClure et al. 2006).

5.5.7.2 Testing procedure

Subjects elevated their arms as far as pain allowed, at which point a ROM reading was taken (see page 414), defined as 'limit of ROM'. The contralateral side (for subjects with unilateral symptoms this was the non-involved side) was then measured and the subject instructed to return to a neutral position. Subjects then performed bilateral elevation again, but this time to inform MS as soon as their symptoms started. For those patients with pain at rest they were instructed to inform MS of the point in range where their symptoms started to get worse. For all subjects this elevation position was recorded and defined as the 'point of symptom exacerbation by movement'. All measurements were taken by the same rater (MS) using a digital inclinometer placed on the mid-point of the lateral aspect of the subject's upper arm, parallel to the humerus (Kolber et al. 2011).

Data processing was as per the pre-methods chapter. However readings were calculated for both 'limit of ROM' (symptomatic side and asymptomatic side) and the 'point of symptom exacerbation by movement' on the symptomatic side.

5.5.8 D3. Clinical measures: Scapular movement and control measures:

5.5.8.1 Recording of standardised scapulo-humeral movement

Subjects were required to expose their scapulae, upper arms and mid/upper back. Female subjects retained their brassieres and all subjects were offered a single-use cotton apron to cover their anterior chest if preferred. As with the ROM measures, subjects stood to perform scaption.

When performing scaption, subjects looked straight ahead with their elbows extended and their thumbs pointing upwards. This was to standardise cervical posture, plane of movement and glenohumeral internal and external rotation (Borstad and Ludewig 2002). Speed of movement was regulated via a metronome (Puretone Digital Metronome) set at 60 beats per minute (1 per second);

4 beats to elevate and 4 to return to resting position i.e. 45 degrees per second. Where subjects had less than full range of movement in an arm they were instructed to move that limb at the same velocity through their available range. Subjects performed bilateral scaption in order to reduce potential compensatory thoracic side flexion (Borstad and Ludewig 2002). Subjects practiced the correct plane and velocity of movement with a sufficient rest period provided prior to the data collection recording.

A digital video camera (Sony Handycam CMOS) was positioned 2 metres behind the subject and the field of view adjusted to capture from waist to above head. It was mounted on a tripod adjusted to the vertical height of the subject's scapulae to ensure a true posterior view with no superior or inferior angulation. Recording occurred in a double clinical cubicle with standardised lighting. This was ensured by minimising natural light and positioning the subject to ensure symmetrical halogen strip lighting was directly above the subject. This allowed subtle shadowing of the scapulae to optimise landmark identification.

A piece of paper with the subject's project code, date and trial type was placed in front of the lens when recording commenced so as to ensure both patient anonymity and project-specific identification of the recording. Subjects performed four, through range scaption movements and then recording was stopped. Subjects were then offered to perform an additional, loaded trial using hand-held weights. Patients self-determined whether they were able to perform this, based upon pain levels and any symptom exacerbation during the non-loaded trial. Patients also self-determined the magnitude of any weight to be used (0.5, 1, 1.5kg), but load symmetry was required.

Subjects who chose to perform a loaded trial practiced the correct plane and velocity of movement with a sufficient rest period prior to the data collection recording. Movement performance and recording was identical to the non-loaded trial except for the piece of paper placed in front of the lens when recording commenced which identified it as a loaded trial. Order of non-loaded versus loaded trial was not randomised due to the potential for disruptive symptom exacerbation in subjects who could not tolerate a loaded trial.

5.5.8.2 Measurement tools

The Scapular Dyskinesis test (SDT) (McClure et al. 2009) was used although the original test involved bilateral shoulder flexion and abduction holding 1.4kg or 2.3kg weights and subsequent qualitative grading. The SDT assessment criteria are recorded in appendix 5.

The McClure et al. (2009) ratings involve a judgement across loaded-only trials involving both flexion and abduction movements. This contrasts with the current study where a single plane of movement was used to make the judgement. However this single composite movement (scaption) was selected due to the emphasis on minimising the exacerbation of symptoms that was likely to occur when multiple movements were performed. Consequently in the current study there is a fundamental deviation from the criteria reported by McClure et al. (2009).

As with the inter-rater reliability study, the analysis of scapular function on the symptomatic side for the cohort study was performed on video recordings of patient movement patterns rather than live assessment. This had the advantage of limiting within and between-trial variability caused by natural variation in patients' movement patterns.

For each subject the grading of their un-loaded and (if performed) loaded trial was recorded for descriptive analysis.

5.5.9 E1. Structural pathology via imaging

Sonographic evidence of structural pathology was assessed at baseline for all prognostic cohort study subjects using the same standardised protocol, differential sonographic diagnoses and proforma for data collection as detailed in the pre-methods chapter.

In light of the reproducibility study findings, an additional mechanism was employed with the aim of increasing the robustness of the sonographic findings, particularly tendinopathy and bursitis. This stemmed from verbal feedback from the experienced musculoskeletal ultrasound practitioner (PM) that his judgements on such sonographic diagnoses were often based on experience derived from the large number of scans he had previously performed.

To overcome this, MS therefore used a composite approach of (i) performing bilateral scans to provide a pseudo comparator and (ii) applied specific criteria relating to the sonographic appearance of cuff tendinopathy and bursitis (appendix 5). Whilst formal assessment of the reproducibility of this approach was not undertaken, this was deemed a constructive attempt to increase the robustness of the sonographic diagnoses in the main cohort (Ingwersen et al. 2016).

All scanning was performed on a Philips CX-50 (Koninklijke Philips N.V.) (appendix 5) using a L12-3 linear array probe with musculoskeletal imaging pre-sets. This portable unit was purchased specifically for the cohort data collection with the expert input of PM. It was in part selected due to its comparability with the Toshiba Aplio units used in the UHW Radiology department for imaging of the shoulder.

As with the ultrasound inter-rater reliability study, findings for each component of the rotator cuff were collapsed to provide a single sonographic diagnosis judgement in relation to the presence of calcific deposits, tendinopathic change, bursitis, partial and full thickness tears. Where one or more components of pathology was identified in the cuff (independent of bursal findings) then a 'cuff pathology' diagnosis was assigned.

5.5.10 E2. Structural pathology via orthopaedic tests

Orthopaedic tests were performed by MS in a standardised manner (Hanchard et al. 2013) for all subjects at baseline; namely Neer's sign, the Empty can and Hawkins-Kennedy test, painful arc and pain on active abduction.

Neer's sign (see page 418) involved MS applying a downward, stabilising force on the subject's scapula whilst simultaneously elevating the arm into full flexion (Michener et al. 2009). Unlike the original Neer's sign (Neer 1983) this was performed without a diagnostic injection (Hughes 2011). A positive finding was defined as replication or worsening of the subject's symptoms (Michener et al. 2009).

The Empty can test (see page 418) involved the subject elevating their arm to 90° flexion in the scapular plane. Subjects were then instructed to internally rotate at the GHJ such that their thumb pointed to the floor and asked to resist downward pressure applied by MS at the distal ulnar (Michener et al. 2009). A positive finding was defined as replication or worsening of the subject's symptoms (Holtby and Razmjou 2004).

The Hawkins-Kennedy test (see page 418) involved MS flexing the shoulder to 90° and internally rotating the humerus with gentle overpressure (Park et al. 2005). A positive finding was defined as replication or worsening of the subject's symptoms (Hughes 2011).

For a positive painful arc finding to be assigned the subject needed to report replication of their pain only between 60° and 120° of active glenohumeral abduction (Kessel and Watson 1977). Pain on active abduction was defined as onset or worsening of their pain during or at the end of the physiological movement.

5.6 Treatment / intervention

Although the nature of the intervention was not considered as a potential prognostic factor, it was nonetheless necessary to record what treatment was delivered so that comparison with other prognostic studies could be meaningfully performed and the clinical transferability established for the findings. As treatment was pragmatic and determined by clinical need, the treatment recording tool detailed in the pre-methods chapter was used.

A training session was arranged at each clinical site where MS presented the treatment collection form to the treating clinicians. In addition, a link clinician was present at each site to provide support with any queries. This form included pages printed up to the 8th appointment with the option for clinicians to request additional pages if required. The form was inserted into the clinical notes and when the patient was discharged, the treatment form was returned to MS. If any missing information was identified then the treating clinician was asked to transcribe onto the form the relevant missing information.

Data processing: the treatment data was entered onto an excel spreadsheet from which the treatment categories delivered per appointment were collated. In addition, details regarding the grade of treating clinician and number of weeks from first to last appointment were recorded.

5.7 How outcome is defined and measured

Outcome was measured using the OSS (Dawson et al. 1996), a validated and reliable, shoulder-specific self-reported measure of function and pain. The OSS was measured at baseline and at two further time points, namely upon discharge from physiotherapy and 3 months post-discharge from physiotherapy. Due to the pragmatic nature of the study, no limit was placed upon the treatment period or discharge time point; instead these were determined by individual clinical need.

The OSS can be seen in appendix 5 ‘OSS: Questionnaire’ and was posted to patients for completion immediately prior to their first appointment. At the first appointment it was placed by MS into a sealed envelope and this procedure was repeated for follow up completions of the OSS. Only once

all variables within the study had been processed and graded did MS open the OSS envelopes, thereby ensuring blinding of MS to the primary outcome measure. The OSS was scored as per appendix 5 'OSS: Scoring' whereby scores range from zero to 48 with 48 being the best outcome.

The treating clinician gave each subject their discharge OSS to complete at the end of treatment. However where patients failed to attend their last appointment MS posted the OSS with a self-addressed, stamped envelope for return. This procedure was repeated for the 3-month follow up and subjects were prompted to return the OSS if not received by MS.

The OSS MCIC of five (Ekeberg et al. 2010b) was used to calculate whether a patient had improved or not at each time point. Specifically an increase in OSS that was greater than or equal to the MCIC denoted 'improved', an increase in OSS that was less than the MCIC or unchanged or a reduction in OSS that was less than the MCIC was denoted 'same'. A reduction that was greater than or equal to the MCIC denoted 'worsened'. This data provided the opportunity to explore the clinical change over the period of treatment and 3 months post completion of treatment.

However as noted in the literature review, outcome defined by a continuous variable retains maximal statistical power (Altman and Royston 2006). Therefore a change score was calculated by taking the OSS scores at discharge from physiotherapy and OSS scores at 3 months post-discharge away from the baseline OSS score for that patient. The dependent variable for the prognostic model was therefore a continuous variable reflecting numerical change in OSS from baseline.

5.8 Prognostic variable trimming approach

The importance of aligning the number of candidate variables with the recruited sample in order to avoid over-fitting was previously highlighted. In light of the sample size requirement of 100 and identification of five categories comprising a total of 10 variables, a systematic approach to variable trimming was required. The following principles were applied:

- Emphasis on logical, theoretical and statistical procedures for trimming the number of candidate variables to align with 10 cases per variable (Peduzzi et al. 1996).
- Retention of minimum of 1 variable per category
- Where novel measurement tool or approaches were used (namely the scapular dyskinesis test and diagnostic ultrasound) then these variables to be retained for consideration in the final regression model.

In addition:

- For prognostic modelling, use of continuous variable to allow for linear regression and retention of maximal statistical power (Altman and Royston 2006).

5.8.1 Sequential decision making process

The approach to trimming the number of candidate variables based upon the principles of logic, theoretical and statistical methods (based on Kromer et al. (2014) was as follows:

1. Clinical relevance; a variable was discarded if its clinical relevance was low compared to other similar variables
2. Conceptual relevance:
    a) where a high % of missing data occurred for conceptual rather than measurement error reasons; a variable was discarded if for conceptual reasons a large proportion of data were likely to be missing which would limit the statistical utility of the variable
    b) Where the incidence of the variable in the cohort was less than 5%; a variable with such low incidence was unlikely to contribute to or influence the subsequent regression model.
    c) where multiple variables can be meaningfully collapsed down into a single score; this provides a mechanism for reducing the number of candidate variables whilst retaining conceptual impact.
    d) where there is high correlation between variables within a category indicating that a single measure could be representative; this provides a mechanism for reducing the number of candidate variables whilst retaining conceptual and statistical impact
    e) where the above rules are of limited utility but the literature indicates that a particular variable is commonly predictive, particularly in relation to alignment with the dependent variable in the current study

The above process was performed once the size of the final recruited sample was known and was further informed by descriptive and univariate analysis of the data.

5.9 Prognostic data analysis approach

Once the candidate variables were identified, they were entered into a multivariate stepwise regression for each model, namely baseline to discharge; and baseline to 3 months post-discharge. SPSS (IBM) version 20 was used. A process of forward selection and backward elimination was applied based on entirely statistical methods, whereby for each variable probability of F to enter ≤ 0.050 and probability of F to remove ≥ 0.100 (Field 2009).

In the subsequent, most parsimonious model, R^2 values were presented to indicate the amount of variation in the change in OSS between the time points explained by each variable and overall model (Bowerman and O'Connell 2000). The adjusted R^2 value will be reported to indicate how well the model generalises from the sample to a population. Analysis of variance (ANOVA) will also be performed to test whether the mean or the model is significantly better as a predictor (Bowerman and O'Connell 2000).

The multiple regression equation coefficients (β) will be used to identify the direction (+ve or –ve) of any relationships between predictor variables and the dependent variable (Miles and Shevlin 2001). The significance of the β values will be used to identify the magnitude of the contribution of the variable. The p values for the variables not featuring in the final regression models will also be stated.

Two fundamental requirements of multiple regression are (a) the dependent variable is continuous and (b) there are two or more independent variables, which can be either continuous (i.e. an interval or ratio variable) or categorical (i.e. an ordinal or nominal variable) (Bowerman and O'Connell 2000). These were met, as per the previous description of the dependent variable (change in OSS score) and the number and nature of the potential prognostic variables.

In addition, appropriate tests of the fit of the model and diagnostics will be performed (Miles and Shevlin 2001; Field 2009):

- Independence of observations will be checked using the Durbin-Watson statistic
- Linearity and homoscedasticity of the residuals will be visually checked by means of a scatterplot of the standardised residuals (errors) against the standardised predicted values
- Absence of collinearity in the data will be checked via the collinearity statistics of the Variance inflation factor (VIF) and Tolerance
- Normal distribution of the residuals will be checked by histogram and PP plots

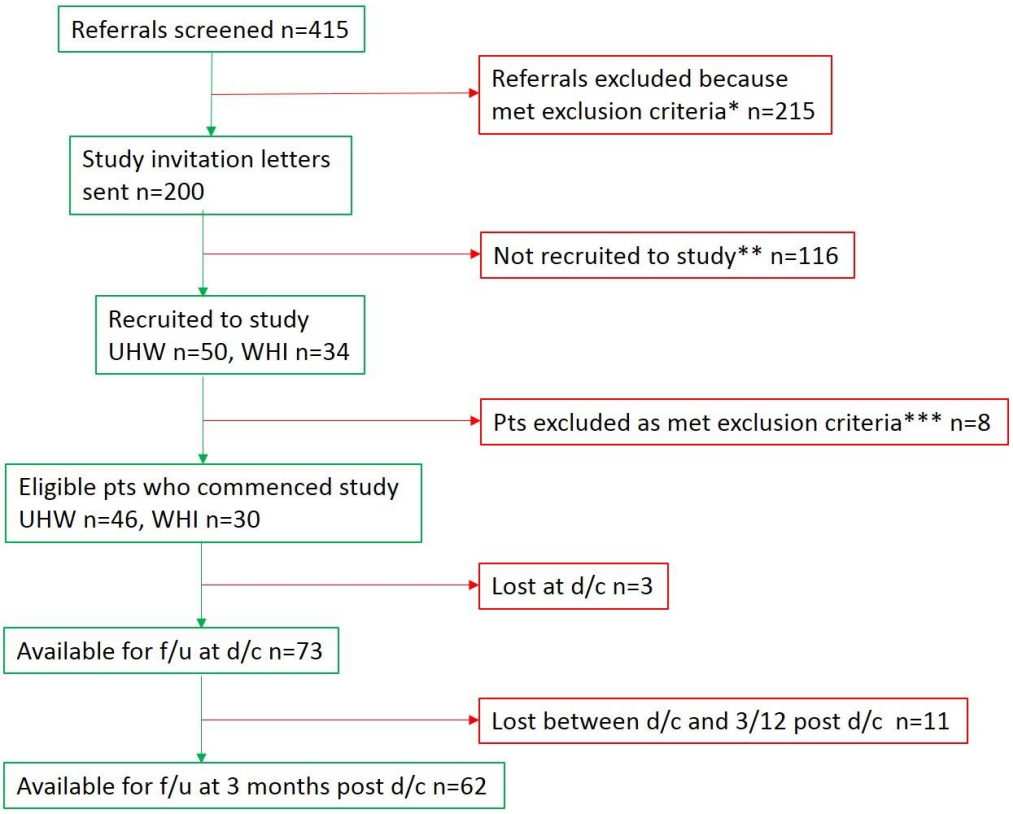
6 RESULTS

This chapter presents the results pertaining to the prognostic cohort study. The nature of the patients recruited to the study will be presented, along with consideration of any statistical differences between those who did and did not consent to participate in the study. The baseline characteristics of those eligible and consented to the study will then be presented and analysed descriptively. The nature of the pragmatic treatment delivered and the clinical outcome as defined by the OSS MCIC will then be considered along with descriptive analysis of the OSS change score between each of the time points. In light of the data characteristics, the data reduction procedures employed for each prognostic variable will be detailed, leading onto an analysis of any difference between those lost to follow up and those for whom follow up was data was available. Finally the regression analysis for each prognostic model will be presented.

6.1 Flow chart of patient recruitment and follow up

Between 19/9/2013 and 23/1/2015 patient recruitment and follow up was as follows:

Figure 6-1: Patient recruitment and follow up



Key: d/c = discharge; f/u = follow up

* Excluded via exclusion criteria (n=):

Post-surgical referral = 5; Trauma leading to a sudden onset of symptoms within the last 4 weeks = 7; Imaging confirmed massive tear = 13; Post-fracture = 20; Cervical spine as the source of the symptoms = 46; Rheumatoid arthritis = 2; Shoulder instability presentation = 26; Adhesive Capsulitis presentation = 49; ACJ as source of symptoms = 13; GHJ osteoarthritis = 18; multi-area symptoms = 16.

** Not recruited to study because (n=):

Declined OR had already made appointment when received study invitation OR couldn't attend on days when study occurring = 77

Patient didn't contact clinical department for treatment and so discharged = 30

Following contact with patient, they did not match the inclusion criteria: (i) underwent surgery between referral and study contact = 2, (ii) symptoms characteristic of cervical referral = 5

Following contact with patient symptoms had resolved so declined appointment = 2

*** Excluded due to exclusion criteria (n=):

Baseline function (as measured by the primary outcome measure of OSS) was within the MCIC of maximum = 4

Treating clinician identified recruited subject as having non-SIS/RCTendinopathy pathology = 2 (adhesive capsulitis), 2 (Cervical spine as the source of the symptoms)

6.2 Sample size

As per the above flow chart, 215 of the 415 patient referrals screened (stage 3 of identification and recruitment of patient sample; methods chapter) were excluded as they met the exclusion criteria. Of these, adhesive capsulitis, cervical referral of the symptoms and shoulder instability were the most common reasons for exclusion. Of the remaining 200 patients, 116 were not recruited to the study. In the majority of cases this was either because of being unable to participate in the study or they declined to participate. However a large number of patients did not contact the department for their clinical appointment and so were directly discharged from physiotherapy. Eighty-four subjects were subsequently recruited to the study. This was less than the 100 subjects stipulated in the sample size, but for time pressures reasons, recruitment had to cease in January 2015.

6.2.1 Characteristics of those who did and who did not consent to participate in the study

Differences between the two groups were assessed to look for any systematic bias. As a continuous variable, age was found to be non-parametrically distributed and so analysed via the Mann-Whitney U test (Altman 1991).

6.2.2 Age

Table 6-1: Age

Variable	Mean	Standard deviation	Range (minimum – maximum)
Consented (n=84)	49.6	13.4	18 - 71
Not Consented (n=116)	51.5	14.7	20 - 83

Table 6-2: Mann-Whitney U result

Mann-Whitney U	Wilcoxon W	Z	Significance
3886.5	6812.5	- 0.912	0.362

As shown in the above table, there was no statistically significant difference between those that consented and those that did not in regard to age.

6.2.3 Gender

As a dichotomous variable, differences in gender was analysed via the chi-square test (Altman 1991).

Table 6-3: Gender

Variable	Male	Female
Consented (n=84)	n=35; 41.7%	n=49; 58.3%
Not Consented (n=116)	n=63; 54.3%	n=53; 45.6%

Chi-square value of 3.117 with p=0.077 indicated no statistically significant difference in gender between those that consented and those that did not.

6.2.4 Source of referral

Table 6-4: Source of referral

Variable	Not recorded	GP	Orthopaedics	Rheumatology	Other
Consented (n=84)	0	n=75; 89.2%	n=6; 7.1%	n=2; 2.4%	n=1; 1.2%*
Not Consented (n=116)	25	n=86; 94.5%^	n=3; 3.3%^	n=0; 0%^	n=2; 2.2%^**

Key: ^ = % calculated relative to the total number recorded for that group, * = 1 x Extended nurse practitioner via A&E, ** = 2 x GP with specialist interest

In both groups the overwhelming majority of patients were referred by their GP. However the large number of patients for whom a referral source was not recorded in those who did not consent introduces uncertainty as to how accurate the percentage values are.

Of the remaining 84 subjects, four were subsequently deemed by their treating clinician, during their treatment phase, as having non-SIS/RCTendinopathy pathology. Specifically two were deemed to have cervical referral of their symptoms and for two the dominant pathology was adhesive capsulitis. A further four patients were subsequently excluded because once all data collection was completed they were identified as having a baseline OSS of 44 or higher. Specifically their baseline OSS were 44 (TM5024h, M0668w) and 48 (M1586h, F8145w). This left n=76 patients as being eligible and consented.

6.3 Characteristics of those eligible and consented to the study (n=76)

6.3.1 A. Demographics

Raw demographic data for the prognostic cohort at baseline is shown in appendix 6 ‘Demographics’ and summarised below.

6.3.1.1 Generic demographic variables: continuous data

Table 6-5: Generic demographic variables: continuous data

Variable	Mean	Standard deviation	Range (min – max)
Age (years)	49.3	13.6	18 - 71
BMI (kgm^{-2})	29.0	6.2	17.3 - 50.5

Key: kgm^{-2} = mass in kilogrammes divided by height in meters squared

6.3.1.2 Generic demographic variables: categorical data

Table 6-6: Generic demographic variables: categorical data

Variable		
Gender	Male: n=31; 40.8%	Female: n=45; 59.2%
Education level	Educated to college or university level: n=62; 81.6%	Less than 12 years in education: n=14; 18.4%

6.3.1.3 Generic demographic variables: ordinal data

Table 6-7: Generic demographic variables: ordinal data

Smoking history	n=	% of cohort at baseline
Current	7	9.2
Previous	19	25.0
Never	48	63.2
Not recorded	2	2.6

The above tables demonstrate a wide range of age and BMI amongst the cohort. They also show a greater proportion of females and a very high incidence of college or university level education amongst the prognostic study cohort. In addition, the majority of subjects had never smoked and less than 10% of the cohort were current smokers.

6.3.2 Work, recreational or functional task related variables

6.3.2.1 Currently working; patient reported

Table 6-8: Currently working; patient reported

Currently working status	n=	% of cohort at baseline
Full time	36	47.4
Part time	13	17.1
Retired	13	17.1
Full time carer	2	2.6
Part time carer	0	0
Student or unemployed	9	11.8
Combination	2	2.6
Not recorded	1	1.3

The above table shows that almost half of subjects were working full time with equal numbers of subjects working part-time as retired. The next largest group were not in paid employment as they were unemployed or were a student. The two subjects with a combination were both working part-time and also studying.

6.3.2.2 Paid work type; patient reported

Table 6-9: Paid work type; patient reported

Paid work type	n=	% of cohort at baseline
Not paid working	23	30.3
Professional / managerial / office	42	55.3
Manual / semi-manual / unskilled	8	10.5
Combination	2	2.6
Not recorded	1	1.3

The above table shows that of those patients who were working, the majority reported they worked in jobs considered professional, managerial or office based. A much smaller number reported they worked in jobs considered manual, semi-manual or unskilled.

6.3.2.3 How often carry 10Kg at work; patient reported

Table 6-10: How often carry 10Kg at work; patient reported

How often carry 10Kg at work?	n=	% of cohort at baseline
Seldom / never	36	47.4
Sometimes	13	17.1
Extremely / often	8	10.5
Not applicable	19	25.0

The above table shows that almost half of the patients identified as seldom or never carrying 10kg at work whilst a smaller number identified as sometimes and only 10% of the cohort identified as extremely or often carrying 10kg. The number of 'not applicable' subjects in the above table was less than the number of subjects who were not in paid working because responses to the question "How often carry 10Kg at work?" also applied to those who were carers.

6.3.2.4 How often work above shoulder height; patient reported

Table 6-11: How often work above shoulder height; patient reported

How often work above shoulder height?	n=	% of cohort at baseline
Seldom / never	35	46.1
Sometimes	13	17.1
Extremely / often	9	11.8
Not applicable	19	25.0

The above table shows an almost identical distribution of those who self-identified as working above shoulder height as those who self-identified as carrying 10kg at work.

6.3.2.5 Play overhead sports?

Forty-seven subjects (61.8%) reported that they did play overhead sports whilst the remaining 29 subjects (38.2%) said they did not.

6.3.3 B. Clinical history

Raw clinical history data for the prognostic cohort at baseline is shown in appendix 6 'Clinical history' and summarised below.

6.3.3.1 Current condition related variables

6.3.3.1.1 Source of referral

Table 6-12: Source of referral

GP	Orthopaedics	Rheumatology	Other
n=69; 90.8%	n=5; 6.6%	n=1; 1.3%	n=1; 1.3%*

Key: * = Extended nurse practitioner via A&E

As per the above table, the overwhelming majority of patients were referred by their GP. A much smaller number came from orthopaedics referrals with only one from each of Rheumatology and A&E.

6.3.3.1.2 Duration of current episode

Table 6-13: Duration of current episode

Duration current episode	n=	% of cohort at baseline
0 to <3 months	6	7.9
3 to <6 months	24	31.6
6 to <12 months	26	34.2
12 to <24 months	8	10.5
24 months or longer	12	15.8

The above table shows that the majority of subjects reported that they had experienced symptoms from the current episode for between 3 and 12 months although 12 subjects reported symptoms for more than 2 years.

6.3.3.1.3 Current condition related variables: categorical data

Table 6-14: Current condition related variables: categorical data

Variable		
Involvement of dominant side*	Symptomatic side = dominant arm: n=35; 46.1%	Symptomatic side = non-dominant arm: n=40; 52.6%
Previous shoulder pain	Previous shoulder pain: n=15; 19.7%	First episode: n=61; 80.3%
Associated neck pain^	Associated neck pain: n=21; 27.6%	No associated neck pain: n=54; 71.1%

Key: * = 1 subject was ambidextrous, ^ = not recorded from 1 subject

The above table illustrates that there was a fairly even split between those whose symptomatic side was also their dominant arm, versus those for whom it was their non-dominant arm. In the majority of cases it was the patient's first episode of shoulder pain that they were presenting with. For almost ¾ of subjects they reported no associated neck pain.

6.3.3.1.4 Precipitating cause

Table 6-15: Precipitating cause

Precipitating cause	n=	% of cohort at baseline
Unknown / insidious / gradual onset	48	63.2
Injury / trauma	11	14.5
Strain / overuse: unusual activities	6	7.9
Strain / overuse: usual activities	9	11.8
Combination	1	1.3
Not recorded	1	1.3

The above table shows that the majority of subjects reported that the precipitating cause of their symptoms was unknown or insidious / gradual onset. Injury or trauma was the next most frequently identified reason for onset although pooled incidence of strain / overuse was higher. The combination was Injury / trauma with Strain / overuse: unusual activities.

6.3.3.1.5 Neurological symptoms

Table 6-16: Neurological symptoms

Neurological symptoms	n=	% of cohort at baseline
None	48	63.2
Pins and needles / tingling	15	19.7
Distal symptoms	9	11.8
Other	1	1.3
Not recorded	2	2.6

The above table shows that the majority of subjects reported no neurological symptoms. Pins and needles / tingling and distal symptoms accounted for the majority of the remaining subjects with one patient reporting 'cold fingers'.

6.3.3.1.6 Course of condition

Table 6-17: Course of condition

Course of condition	n=	% of cohort at baseline
Worse	21	27.6
Same	40	52.6
Better	14	18.4
Not recorded	1	1.3

The above table shows that just over half of subjects reported their symptoms as being static whilst more reported their symptoms worsening than improving.

6.3.3.2 Treatment related variables

6.3.3.2.1 Previous rehabilitation

Table 6-18: Previous rehabilitation

Previous rehabilitation	n=	% of cohort at baseline
None	58	76.3
Physiotherapy	14	18.4
Chiropractic	2	2.6
Private massage	2	2.6

The above table shows that the majority of subjects reported no previous rehabilitation. Of those that had, most received physiotherapy.

6.3.3.2.2 Current condition related variables: categorical data

Table 6-19: Current condition related variables: categorical data

Variable		
Currently taking analgesia/NSAIDS*	Yes: n=26; 34.2%	No: n=49; 64.5%
This episode; number of GP injections^	1 injection: n=9; 11.8%	2 injections: n=3; 3.9%

Key: * = not recorded for 1 patient, ^ = 64 subjects did not receive an injection from their GP

Just over a third of patients were currently taking oral medication for their pain. Just over 15% of subjects had received an injection via their GP, of whom most had received one injection.

6.3.3.2.3 This episode; treatment via GP

Table 6-20: This episode; treatment via GP

This episode; treatment via GP	n=	% of cohort at baseline
No contact	3	3.9
Advice only	2	2.6
Oral or topical medication (analgesia / NSAIDs) (+/- advice)	27	35.5
Oral medication + injection	6	7.9
Injection only	6	7.9
Referral only	32	42.1

The above table shows that the majority of subjects saw their GP and were either directly referred on to physiotherapy or received analgesia / NSAIDs before being referred on. The same number of subjects had an injection via their GP as those who received oral medication and an injection. Note that the above dataset is for n=76 even though only n=69 were referred to physiotherapy via their GP. This was because all except three patients had contact with their GP at some point in the care pathway for this current episode.

6.3.3.2.4 This episode; injection by non GP

Seventy-four subjects (97.4%) reported that they had not received an injection from someone other than their GP. Of those that did, in both cases it was an orthopaedic surgeon who had administered the shoulder injection. Of these two subjects, one also received two injections via their GP; whilst the other subject was simply referred on by their GP.

6.3.3.2.5 This episode; investigations

Table 6-21: This episode; investigations

This episode; investigations	n=	% of cohort at baseline
None	38	50.0
U/S only	10	13.2
MRI only	0	0
X-ray only	17	22.4
X-ray + U/S	7	9.2
Other combination	3	3.9
Not recorded	1	1.3

Key: U/S = Ultrasound, MRI = Magnetic resonance imaging

The above table shows that exactly half of subjects had no investigations performed. X-ray only followed by ultrasound only and then X-ray + ultrasound were the most undertaken investigations. Other combinations were MRI + X-ray (n=2) and ultrasound + MRI + X-ray (n=1).

6.3.4 C1. Patient reported measures: Pain

Raw pain data for the prognostic cohort at baseline is shown in appendix 6 and summarised below.

Table 6-22: Descriptive analysis of pain data (cm) for the prognostic cohort study subjects

Variable	n =	Measure of variability			
		Mean	SD	Min	Max
'On activity'	74	5.9	2.2	0.2	9.1
'At rest'	74	2.0	2.0	0.0	8.7
'At night over the last week'	76	4.4	2.9	0.0	9.4
Mean of 3 pain scores	74	4.1	1.9	0.5	8.7

Key: SD = standard deviation, Min = minimum, Max = maximum

The above table shows a higher mean pain reading for 'On activity' than 'At night over the last week', with 'At rest' having the lowest mean score. Whilst the SD demonstrate comparable spread of data across all 4 variables, greatest absolute variability is seen with the 'At night over the last week' pain score and greatest variability relative to the mean is seen with the 'At rest' pain score.

The large range from minimum (0.0 to 0.5 across all 4 variables) to maximum (8.7 to 9.4 across all 4 variables) again highlights the large variability within the cohort.

The heterogeneous nature of the cohort is highlighted by subjects such as M2747w with high 'On activity' scores but zero for 'At rest' and 'At night over the last week'; subjects such as M1662h with very high pain scores across all parameters; and subjects such as M8695h with very low pain scores across all parameters.

6.3.5 C2. Patient reported measures: Psychological symptoms

Raw 4DSQ data for the prognostic cohort at baseline is shown in appendix 6 and summarised below.

Table 6-23: Descriptive analysis of 4DSQ data for the prognostic cohort study subjects

Variable	n =	Not elevated (n= ; % of respondents)	Moderately elevated (n= ; % of respondents)	Strongly elevated (n= ; % of respondents)
Distress	74	n=56 ; 75.7%	n=13 ; 17.6%	n=5 ; 6.8%
Depression	74	n=65 ; 87.8%	n=5 ; 6.8%	n=4 ; 5.4%
Anxiety	74	n=63 ; 85.1%	n=8 ; 10.8%	n=3 ; 4.1%
Somatisation	74	n=58 ; 78.4%	n=15 ; 20.3%	n=1 ; 1.4%

4DSQ data was not available for subjects F6847h and M2146h due to non-completion of the questionnaire and there being insufficient time for its completion during the face-to-face assessment period. This gave 4DSQ cohort data for 74 subjects.

The above table shows that there was a low incidence of 'strongly elevated' psychological symptoms in the cohort. However 'moderately elevated' psychological symptoms were more prevalent, particularly those categorised as signifying symptoms of 'distress' and 'somatisation'.

As shown in the raw 4DSQ data (appendix 6 '4DSQ: absolute score') there was wide variation in the levels and pattern of psychological symptoms reported. For example, subjects such as F4417h and M8305w were classified as being below 'moderately elevated' across all categories. Conversely subjects such as M7524w were rated as 'moderately elevated' in the majority of categories whilst other subjects such as F9240h were rated as 'strongly elevated' in the majority of categories. Finally

some subjects such as M0035w were rated as ‘strongly elevated’ in a single category (distress) but below ‘moderately elevated’ across all other categories.

6.3.6 C3. Patient reported measures: Function / Disability

Raw SPADI data for the prognostic cohort at baseline is shown in appendix 6 and summarised below.

Table 6-24: Descriptive analysis of SPADI data (%) for the prognostic cohort study subjects

Variable	n =	Measure of variability			
		Mean	SD	Min	Max
Total pain score	76	53.3	21.9	6.0	100.0
Total disability score	76	36.7	23.3	0.0	95.0
Total SPADI score	76	43.2	21.6	2.3	96.9

Key: SD = standard deviation, Min = minimum, Max = maximum

The above table shows the total pain score had the highest mean value (53.3%) of the three variables with total disability score the lowest mean value (36.7%). The SD for each variable was very comparable (21.6% to 23.3%) and reflected a large variation in ‘total score’ data for each variable. In relation to the mean, this variability was largest for the total disability score variable.

As shown in the raw SPADI data in appendix 6 there was wide variation in individual scores and patterns of response. For example, subjects such as M1662h and F9240h had very high scores across all three variables, subjects such as M8695h had very low scores across all three variables and some subjects such as F6847h had high total pain scores but much lower total disability scores.

6.3.7 D1. Clinical measures: Strength

The raw strength data for the prognostic cohort study subjects is shown in appendix 6 and the descriptive analysis of the strength variables are shown in the below table.

Table 6-25: Descriptive analysis of strength (Nm) data for the prognostic cohort study subjects

Variable		n =	Measure of variability			
			Mean	SD	Min	Max
Internal Rotation	Symptomatic side Mean moment	75	11.5	8.2	2.1	52.7
	Asymptomatic side Mean moment	75	14.3	8.8	4.8	58.0
External Rotation	Symptomatic side Mean moment	75	10.6	6.7	2.1	35.5
	Asymptomatic side Mean moment	75	13.6	7.1	2.2	36.3
Scaption	Symptomatic side Mean moment	75	11.7	9.9	0.0	60.8
	Asymptomatic side Mean moment	75	17.7	10.7	4.1	57.2

Key: SD = standard deviation, Min = minimum, Max = maximum

The above table shows that for each contraction type (IR, ER and scaption) the mean moment on the symptomatic side was less than on the asymptomatic side and this difference was greatest for scaption. Furthermore for each contraction type and side there was substantial variability in the data as demonstrated by the large SD values relative to the respective means.

Strength data were available for 75 subjects because of an error with the unit with subject M2279h. Subject F4417h had a recorded strength value of zero for scaption on the symptomatic side, as this subject was unable to reach the unit's recording threshold (4.4N) due to pain during the test movement. Subject M4091h reported symptoms on the 'asymptomatic side' at the time of testing.

6.3.8 D2. Clinical measures: ROM

The raw ROM data for the prognostic cohort study subjects is shown in appendix 6 and the descriptive analysis of the ROM variables are shown in the below table.

Table 6-26: Descriptive analysis of ROM (°) data for the prognostic cohort study subjects

Variable	n =	Measure of variability			
		Mean	SD	Min	Max
Symptomatic side Limit of ROM	76	131.2	31.8	59.0	178.0
Symptomatic side Point of symptom exacerbation by movement	66	102.2	29.4	34.0	155.0
Asymptomatic side Limit of ROM	76	157.7	12.3	110.0	179.0

Key: SD = standard deviation, Min = minimum, Max = maximum

The above table shows the mean 'limit of ROM' on the symptomatic side (131.2°) with a large SD of 31.8° and range of 59.0° to 178.0°. This reflects a large variation in 'limit of ROM' data on the symptomatic side in the prognostic cohort.

Compared to the 'limit of ROM' data, the table shows the lower mean 'point of symptom exacerbation by movement' ROM on the symptomatic side (102.2°) with again a large SD of 29.4° and range of 34.0° to 155.0°. As with 'limit of ROM', this reflects the large variation in 'point of symptom exacerbation by movement' in the prognostic cohort. However as shown in the raw ROM data (appendix 6), 8 subjects reported no symptoms during movement whilst 2 subjects reported pain which remained the same throughout the ROM that they performed. As such, no 'point of symptom exacerbation by movement' ROM data was entered for these subjects. Of the remaining 66 subjects, 10 reported the 'point of symptom exacerbation by movement' to also be their 'limit of ROM' whilst 56 reported their 'point of symptom exacerbation by movement' as being during their movement towards their 'limit of ROM', i.e. lower in their range than their 'limit of ROM'.

Unsurprisingly the mean 'limit of ROM' on the asymptomatic side (157.7°) was greater than on the symptomatic side; and the spread of data (SD = 12.3°) was less than either ROM variable on the symptomatic side. However the minimum was 110° and the raw ROM data reveals that 40 subjects

had a 'limit of ROM' of less than or equal to 160° on their asymptomatic side. This included the two subjects (M3288h, M4091h) who reported bilateral symptoms at the time of testing. Of the subjects with unilateral symptoms, 50 had a lower 'limit of ROM' on the symptomatic side than the asymptomatic side. However, 18 subjects with unilateral symptoms had the same 'limit of ROM' bilaterally whilst six actually demonstrated a higher 'limit of ROM' on their symptomatic side compared to their asymptomatic side.

6.3.9 D3. Clinical measures: Scapular movement and control

The raw scapular dyskinesis data (SDT) for the prognostic cohort study subjects is shown in appendix 6 and the descriptive analysis of the SDT grading variables are shown in the below table.

Table 6-27: Descriptive analysis of SDT data for the prognostic cohort study subjects

Trial type (n=)	SDT Grading (n=; %)		
	Normal	Subtle abnormality	Obvious abnormality
Un-loaded (76)	26; 34%	31; 41%	19; 25%
Loaded (45)	9; 20%	28; 62%	8; 18%

The above table shows that all subjects completed an un-loaded trial and there was a fairly even spread in the grading of normal, subtle and obvious abnormality across the cohort. Thirty-one subjects (42% of cohort at baseline) did not perform a loaded trial. Where a loaded trial was performed, subtle abnormality was the most commonly applied grading. From the raw scapular dyskinesis data it can be seen that 0.5kg was the most commonly used weight (n=32), then 1.0kg (n=13) with no subjects using 1.5kg. Of those who performed a loaded trial (n=45) the grading was unchanged between the un-loaded and loaded trial in the majority (n=36) of subjects. For seven subjects the grading increased by one grade whilst for two subjects it decreased by one grade.

6.3.10 E1. Structural pathology via imaging

The raw US data for the prognostic cohort study subjects are shown in appendix 6 along with examples of how bilateral scanning was used as a mechanism to assist with tendinopathic differential diagnoses. Descriptive analysis of the sonographic diagnosis variables are shown in the below table.

Table 6-28: Descriptive analysis of US data for the prognostic cohort study subjects

Variable	n =	Yes	No
Cuff pathology	76	n=46 ; 60.5%	n=30 ; 39.5%
Tendinopathy	76	n=44 ; 57.9%	n=32 ; 42.1%
Calcific deposits	76	n=6 ; 7.9%	n=70 ; 92.1%
Bursitis	75	n=38 ; 50.7%	n=37 ; 49.3%
PTT	76	n=2 ; 2.6%	n=74 ; 97.4%
FTT	76	n=3 ; 3.9%	n=73 ; 96.1%

Key: PTT = partial thickness tear; FTT = full thickness tear

The above table shows that approximately half of subjects had sonographic evidence of bursitis and almost two-thirds had sonographic evidence of cuff pathology. There was a low incidence of subjects with sonographic evidence of calcific deposits (n=6), PTT (n=2) and FTT (n=3). Sonographic evidence of tendinopathy was a common finding (n=44).

6.3.11 E2. Structural pathology via orthopaedic tests

Data on five orthopaedic test findings (Neer's sign, the Empty can and Hawkins-Kennedy test, painful arc and pain on active abduction) was collected.

However in relation to the painful arc data, 2 subjects (M9231w and M3960h) reported pain that remained the same throughout their range of movement. Furthermore 26 subjects (F4417h, F4583h, F2520h, F1005w, F4764h, F0809h, F8774w, F2387h, F8396w, F4674w, M1243h, F8966w, M1662h, F4142w, F9240h, F9830h, M2279h, M1543h, F1188h, F5706h, F9261h, M7524w, F3795w, F8809w, F6914w and M3960h) did not have the ability to reach 120° of active glenohumeral abduction (appendix 6 ‘D2. Clinical measures: ROM’) which was the ROM necessary for a painful arc finding to be considered (Kessel and Watson 1977).

In addition, the challenge of patients accurately identifying the point in the range of movement where symptom onset began and the range of movement where symptoms ceased was complicated by the overlap between symptom onset or worsening and symptoms ceasing or easing. These conceptual and measurement issues threatened the robustness of the variable and so the painful arc data was not analysed. Consequently, the raw structural pathology via orthopaedic tests data for four tests recorded from the prognostic cohort study subjects is shown appendix 6.

Table 6-29: Descriptive analysis of orthopaedic test data for the prognostic cohort study subjects

Variable	n =	+ve	-ve
Hawkins-Kennedy	74	n=50 ; 65.8%	n=26 ; 34.2%
Neer's sign	74	n=59 ; 77.6%	n=17 ; 22.4%
Empty Can	73	n=55 ; 73.3%	n=20 ; 26.7%
Pain on active shoulder elevation	74	n=68 ; 89.5%	n=8 ; 10.5%

As can be seen from the above table there was a high incidence of positive test findings ranging from 66% of the cohort having a positive result with Hawkins-Kennedy test through to 78% for Neer's sign. The one subject (F6847h) with no empty can finding was unable to achieve the test position due to pain. Only 11% of the cohort did not have pain on active abduction.

6.4 Treatment / intervention

The below table presents the grade of the treating clinician and the treatment delivered to each patient per appointment.

Table 6-30: Grade of the treating clinician and the treatment delivered to each patient per appointment

Patient code	Clinician grade	Appt 1	Appt 2	Appt 3	Appt 4	Appt 5	Appt 6	Appt 7	Appt 8	Appt 9
F6847h	6	C, G	-	-	-	-	-	-	-	-
F7047h	7	A, C, G	A, B, C, G	A, B, C, G	A, B, C, G	A, C, G	-	-	-	-
M2146h	6	A, C, G	A, C, G	-	-	-	-	-	-	-
F5446h	6	A, C, G	A, C, G	A, C, G	A, C, G	A, C, G	A, C, G	C, G	A	-
F4417h	6	A, C, G	C, G	C, G	C, G	F	F	F	F	F*
M0379h	6	A, B, C, F, G	A, B, C, G	A, B, C, G	A, C, G	C, G	-	-	-	-
F4583h	6	A	A, C, E, G	C, E, G	A, C, E, G	F	A	A, C, G	-	-
F7849h	6	A, C, G	A, C, G	A, C, G	A, C, G	-	-	-	-	-
M0825h	6	A, B, C, G	A, B, C, G	A, B, C, G	A, C, G	A, G	-	-	-	-
M0061h	6	A, B, C	A, B, C, G	A, B, C, G	A, C, G	A	A, C, G	A, B, C, G	A, C, G	-
F2520h	6	A, B, C	A, C, G	A, B, C, G	B, C, G	G	-	-	-	-
F8486h	7	A, B, C, G	A, B, G	A, B, C, G	B, G	A, C, G	-	-	-	-
F1005w	6	A, C, G	C, G	A	A, C, G	A, C, G	-	-	-	-
F6416h	6	A, B, C, G	A, B, C, G	-	-	-	-	-	-	-
F4764h	6	A, B, C, G	A, B, C, G	A, C, G	B, C, G	C, G	A, C, G	-	-	-
M4542h	6	A, B, C, G	A, B, C, G	B, C, G	A, B, C, D	B, C, G	C, G	C, G	A, C, G	A, G

M8878h	6	A, C, G	A, B, C, G	C, G	A, G	-	-	-	-	-
F0165h	6	A, C, G	A, C, G	C, G	C, G	C, G	-	-	-	-
F5367w	6	A, C, G	B	A, B, C, G	A	-	-	-	-	-
F7405h	6	A, B, C, G	B	B, G	No Rx	-	-	-	-	-
M2747w	7	A, C, G	A, C, G	A, C, G	A	-	-	-	-	-
F1634w	6	A, C, G	A, C, G	C, G	C, D, G	A, C, G	A, C, G	C, G	A, C	-
F0738w	6	C, G	-	-	-	-	-	-	-	-
M0035w	7	A, C, G	C, G	-	-	-	-	-	-	-
F5503w	6	A, B, C, G	A, C, G	A, C, G	A	-	-	-	-	-
M9819h	8	A, B, C, G	A, B, C, G	-	-	-	-	-	-	-
M7535w	6	A, G	A, D, G	A, C, G	C, G	A, C, G	A, C, G	C, G	No Rx	-
F0809h	7	A, C, G	A, C, G	A, C, G	A, C, G	-	-	-	-	-
F4113w	6	C, G	B, C, G	C, G	B, C, G	G	C	-	-	-
F9939h	7	A, B, C, G	B, C, G	B, G	B, G	B	-	-	-	-
M5703h	7	A, B, C, G	B, C, G	A, B, C, G	A, C, D, G	-	-	-	-	-
M1518h	6	A, C, G	C, G	C, G	C, G	A, C, G	C, G	C, G	-	-
F8774w	6	A, B, C, G	A, B, C, G	-	-	-	-	-	-	-
M8695h	7	A, C, G	A, C, G	-	-	-	-	-	-	-
F7304h	7	A, C, G	A, B, C, G	A, B, C, G	A, B, C, G	-	-	-	-	-
F2387h	6	A, C, G	A, C, G	C, G	-	-	-	-	-	-
F8396w	6	A, B, C, G	B, C, G	B, C, G	C, G	A, C, G	C, G	C, G	C, G	C*
F4674w	6	A, B, C, G	A, B, C, G	A, B, C, G	B, C	A	C, G	A, G	-	-

M1243h	5	A, B, G	G	A, C, G	B, C, G	A, G	-	-	-	-
M2593	6	A, C, G	A, C, G	-	-	-	-	-	-	-
M3288h	6	A, C, G	B, C, G	-	-	-	-	-	-	-
M0519h	6	A, C, G	A, C	-	-	-	-	-	-	-
M2608w	6	A, C, G	A, C, G	A, C, G	A, C, G	C, G	C, D, G	-	-	-
F8966w	6	A, C, G	C, G	C, G	A, C, G	A, C, G	-	-	-	-
F4943h	7	A, C, G	A, C, G	A, C, G	-	-	-	-	-	-
F5204w	6	A, B, C, G	C, G	C, G	C, G	C, G	C, G	-	-	-
M1662h	7	A, C, G	-	-	-	-	-	-	-	-
M5465w	6	A, B, C, G	C, G	A	-	-	-	-	-	-
M8305w	6	A, C, G	A, B, C, G	B, C, G	B, C, G	B, C, G	B, C, G	C, G	B, C, G	B, C, G*
F0732w	7	A, C, G	A, B, C, G	B, C, G	C, G	C, G	-	-	-	-
F2122w	6	A, C, G	C, G	-	-	-	-	-	-	-
M2028w	7	n/r	-	-	-	-	-	-	-	-
F4142w	6	A, C, G	B, C, G	A, C, G	C, G	A, C, G	B, C, G	A, B, C, G	-	-
F9240h	6	A, F	A, C, F, G	A, C, G	A, B, C, G	A, B, C, G	A, G	A, G	A	-
F2537h	5	A, C, G	A, C, G	A, C, G	G	-	-	-	-	-
F4735h	5	Ax only	A	A, C, F, G	A, C, G	A, C, G	-	-	-	-
F1965w	7	A, C, G	-	-	-	-	-	-	-	-
M6858w	6	A, C, G	-	-	-	-	-	-	-	-
F9830h	6	A, C, G	A, B, G	-	-	-	-	-	-	-
M2279h	6	A, B, C, G	B, C, G	B, C, G	B, G	No Rx	-	-	-	-

M1543h	6	A, B, C, G	B, C, G	No Rx	-	-	-	-	-	-
F1188h	7	A, C, G	-	-	-	-	-	-	-	-
M4091h	5	A, G	A, C, G	A, C, G	A, C, G	-	-	-	-	-
F5706h	7	A, C, G	B, C, G	A, C, D, G	B, C, G	A, B, C, F, G	A, C, G	A	-	-
F1669h	6	A, C, G	A, C, G	A	-	-	-	-	-	-
F9261h	6	A, C, G	C, G	A, B, C, G	B, C, G	C, G	A, C, G	-	-	-
M7524w	6	A, C, G	A, C, G	C, G	-	-	-	-	-	-
F3795w	6	A, B, C, G	A, C, G	C, G	C, G	B, C, G	C, G	C, G	C, G	C, G+
M9231w	6	Ax only	A, C, G	A, C, G	A, C, G	A, C, G	C, G	A, C, G	-	-
F5690h	6	A, B, C	C	C, G	No Rx	-	-	-	-	-
M3609h	5	A, G	-	-	-	-	-	-	-	-
F8809w	6	A, C, G	B, C, G	A, B, C, G	A, B, C, G	-	-	-	-	-
F6914w	6	A, C, G	C, G	C, G	B, C, G	B	-	-	-	-
M3960h	6	A, C, G	A, B, C, G	A, C, G	A, C, G	-	-	-	-	-
M9287h	7	A, B, C, G	B, C, G	-	-	-	-	-	-	-
F2197w	5	A, C, G	A, C, F, G	C	-	-	-	-	-	-

Key: Appt = appointment, Ax only = Assessment only, No Rx = No treatment delivered in appointment

For treatment categories:

A = Education and advice

B = Manual therapy

C = Exercise therapy

D = Taping

E = Electrotherapy

F = Other treatment

G = Home exercise programme

F4417h; F* denotes 1 further hydrotherapy treatment session and then a further appointment to review and discharge

F8396w; C* denotes 1 further appointment but no treatment delivered in the last appointment

M8305w; G* denotes 1 further appointment where home exercise programme was emphasised
M2028w; n/r denotes not recorded. Patient sustained a cerebral vascular accident independent to treatment but during the treatment phase. Treatment notes were not available for inputting.
F3795w; G+ denotes 4 further appointments comprising A, C, G; C, G; C, G; A

The number of patients treated by each grade of clinician were: n=6 (7.9%) by band 5, n=52 (68.4%) by band 6, n=17 (22.4%) by band 7 and n=1 (1.3%) by band 8. Therefore the majority of patients were treated by grade 6 and 7 clinicians.

With regards to the treatment delivered in each appointment, there was a predominance of (A) education and advice (n=72; 95%), (C) exercise therapy (n=74; 99%) and (G) a home exercise programme (HEP) (n=75; 100%). Please note that the percentage of patients is calculated relative to the 75 for whom treatment data was available. Manual therapy was used in the treatment of half (n=38) of patients whilst taping (n=6; 8%) and electrotherapy (n=1; 1%) were used on a much smaller portion of the sample. For (F) other treatments these were hydrotherapy (F4417h), hot packs (M0379h and F9240h), shoulder injection (since physiotherapy treatment commenced) from General Practitioner (F4583h and F5706h), cold packs (F4735h) and acupuncture (F2197w).

The raw data for the number of weeks under treatment and discharge situation is seen in appendix 6. The number of appointments varied widely around a mean of 4.6 and a standard deviation of 2.6. Some notable outliers were the seven patients who each only attended one appointment, which contrasts with patients such as F4417h, F8396w, M8305w and F3795w who had 11, 10, 10 and 13 appointments respectively.

The length of time over which treatment occurred also varied widely around a mean of 13 weeks and standard deviation of 10 weeks, ranging from zero to 42 weeks. Again, this reflects a high degree of heterogeneity within the data and the wide range of treatment circumstances. This included patients such as F0738w, M1662h, F1965w, M6858w and F1188h who only attended for their first appointment before being discharged from physiotherapy due to failure to attend a subsequent appointment. Other patients (F9939h, M5703h, F5690h, and F2197w) attended multiple appointments before being discharged from physiotherapy due to failure to attend a subsequent appointment.

Typically, those patients with a larger number of appointments were also under treatment for the largest period of time. However, notable exceptions are F0732w who attended 5 times over a 32-

week period and F2197w who only attended 3 times over the same time period. Again this reflects the wide variation in treatment circumstances and clinical management.

The majority of patients at discharge (n=57; 75%) had completed physiotherapy treatment and were subsequently discharged. However 13 patients (17%) did not attend (DNA) or were unable to attend (UTA) and were subsequently discharged from physiotherapy. Five patients (7%) were transferred to the care of a different speciality which was orthopaedics for surgery consultation (M9231w and F6914w), injection via orthopaedics (F6847h) or GP (M1543h) and specialist psychological treatment (F1634w). One patient (M2028w) did not complete treatment for medical reasons unrelated to the shoulder.

6.5 Outcome

6.5.1 Descriptive analysis for OSS at each time point

Loss to follow up meant that 73 and 62 patients were available for discharge and 3 months post discharge follow up; respectively. This represented a 96.1% and 81.6% retention rate at the 2 follow up time points relative to those eligible and consented (n=76).

The raw data for the patient reported outcome (OSS) at each time point are shown in appendix 6 'Outcome measure' and summarised below.

Table 6-31: Descriptive analysis of OSS score for the prognostic cohort study subjects

Variable	n =	Measure of variability			
		Mean	SD	Min	Max
Baseline	76	31.6	7.7	8	43
Discharge	73	39.7	8.0	10	48
3 months post-discharge	62	41.0	8.3	14	48

Key: SD = standard deviation, Min = minimum, Max = maximum

The above table illustrates that at baseline, mean OSS was lower than at discharge or at 3 months post-discharge. Absolute variability in the data was largely similar at each time point.

6.5.2 Clinical outcome defined by OSS MCIC

The data for the categorised change in patient reported outcome (OSS) between each time point are shown in appendix 6 'Outcome measure' and summarised below.

Table 6-32: Categorised change using OSS MCIC from baseline

Categorised change	Baseline to d/c (Total n=73)	Baseline to 3 months post d/c (Total n=62)
'Improved'	n=51; 69.9%	n=47; 75.8%
'Same'	n=18; 24.7%	n=12; 15.8%
'Worse'	n=4; 5.4%	n=3; 3.9%

Key: d/c = discharge

The above table demonstrates that the largest proportion of subjects were 'improved' from baseline at both the discharge and 3 months post discharge time points. A very small proportion of subjects worsened between the same time points with the remaining subjects unchanged.

Discharge OSS data were unavailable for M2028w, F2197w and M9819h. This was because M2028w did not complete treatment for medical reasons (stroke) unrelated to the shoulder, whilst the other subjects did not return their discharge OSS forms. Three months post-discharge data were unavailable for M2146h, F1634w, F9939h, M5703h, F2387h, F0732w, M6858w, F1188h, F1669h, F5690h and M3609h as these subjects did not return their 3 months post-discharge OSS forms.

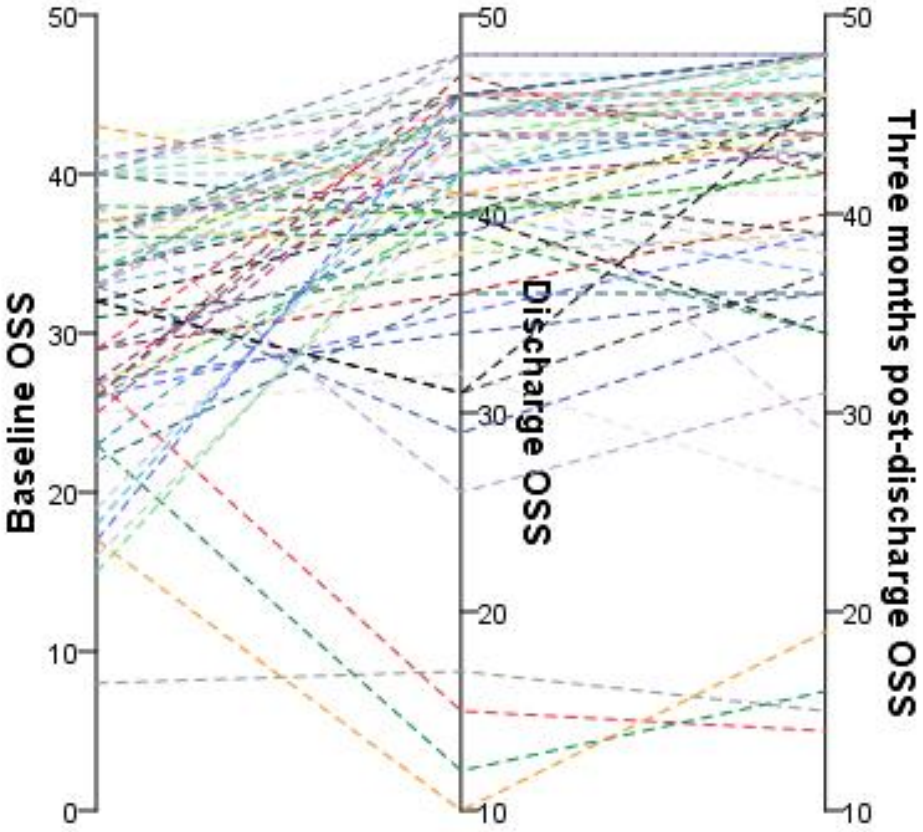
Any systematic differences between those available for follow up at discharge and at 3 months post discharge compared to those lost to follow up will be explored. In order that a focused analysis is undertaken, this will only be performed once the variables to be considered for the final prognostic models have been identified.

Table 6-33: Categorised change using OSS MCIC from baseline to discharge to 3 month post-discharge

Categorised change from Baseline to Discharge to 3 month follow up	Number (%) (Total n=62)
1a = 'Improvement' from baseline to discharge and 'Improvement' maintained at 3 month f/u	n=34 (54.8%)
1b = 'Improvement' from baseline to discharge and further 'Improvement' from discharge to 3 month f/u	n=4 (6.5%)
1c = 'Improvement' from baseline to discharge; 'worsening' (according to MCIC) from discharge to 3 month f/u, but 3 month f/u level still 'Improved' relative to baseline	n=1 (1.6%)
2a = 'Same' from baseline at any time point	n=5 (8.1%)
2a* = 'Same' from baseline at any time point; note baseline level meant that maximum OSS score required to demonstrate 'Improvement'	n=1 (1.6%)
2b = 'Worse' from baseline to discharge and 'Worse' compared to baseline maintained at 3 month f/u	n=3 (4.8%)
2c = At discharge was 'worse' compared to baseline; at 3 month f/u was within MCIC of baseline	n=1 (1.6%)
3 = 'Same' at discharge but at 3 month f/u had 'Improved' from baseline	n=8 (12.9%)
4 = 'Improved' from baseline to discharge but within MCIC of baseline (i.e. Same relative to baseline) at 3 month f/u	n=5 (8.1%)

The above table demonstrates a complex pattern of change across the three time points. The largest proportion of subjects improved from baseline to discharge and maintained this relative to baseline at 3 months post-discharge. The next largest group were those who were the same at discharge but at 3 months post-discharge had improved from baseline. The remaining subjects were fairly evenly split amongst the other permutations.

Figure 6-2: Representation of change in OSS over time for each subject



The above graph highlights the highly variable nature of the individual changes in OSS over time. A small number of very low and dramatic worsening of OSS scores after baseline can be seen. However the majority of subjects improve from baseline to discharge and/or 3 months post discharge.

6.5.3 OSS change score as dependent variables for prognostic modelling

The data for the numerical change score derived from the patient reported outcome (OSS) between baseline and each follow up time point is shown in appendix 6 'Outcome measure' and summarised below.

Table 6-34: Descriptive analysis of OSS change score for the prognostic cohort study subjects

Variable	n =	Measure of variability			
		Mean	SD	Min	Max
Change score from baseline to discharge	73	8.3	8.9	-12.0	29.0
Change score from baseline to 3 months post-discharge	62	9.8	8.6	-13.0	33.0

Key: SD = standard deviation, Min = minimum, Max = maximum

The above table illustrates that the mean change score at each time point was +ve indicating an improvement in pain and function. However the large SD relative to each mean along with the wide range of scores demonstrates a large spread of data for each variable.

6.6 Data reduction

The selection of baseline variables for consideration in the final prognostic model was achieved via the prognostic variable trimming approach detailed in the method section.

6.6.1 A. Demographics

1. Clinical relevance – all of the variables were considered to be potentially clinically relevant

2. Conceptual relevance:

a) where a high % of missing data occurs for conceptual rather than measurement error reasons – this was not applicable

b) where the incidence of the variable in the cohort was less than 5% – this was not applicable

c) where multiple variables can be meaningfully collapsed down into a single score – this was not applicable

d) where there is high correlation between variables within a category indicating that a single measure would be representative – this was not applicable

e) where the above rules are of limited utility but the literature indicates that a particular variable is commonly predictive, particularly in relation to alignment with the dependent variable in the current study.

Age and gender were the two variables most commonly predictive. However age was selected over gender because there is evidence of age predicting change in status (as well as outcome

state), thereby aligning more closely with the handling of the outcome variable in the current study.

6.6.2 B. Clinical history

1. Clinical relevance – all of the variables were considered to be potentially clinically relevant

2. Conceptual relevance:

a) where a high % of missing data occurs for conceptual rather than measurement error reasons – this was not applicable

b) where the incidence of the variable in the cohort was less than 5% – Treatment related variable: injection by non-GP was only 2.6% of the cohort, so this variable was excluded

c) where multiple variables can be meaningfully collapsed down into a single score – this was not applicable

d) where there is high correlation between variables within a category indicating that a single measure would be representative – this was not applicable

e) where the above rules are of limited utility but the literature indicates that a particular variable is commonly predictive, particularly in relation to alignment with the dependent variable in the current study.

Symptom duration was selected because it was overwhelmingly the most commonly identified clinical history predictor in the literature, including four studies that found it to be predictive of change in status.

6.6.3 C1. Patient reported measures: Pain

1. Clinical relevance – all of the variables were considered to be potentially clinically relevant

2. Conceptual relevance:

a) where a high % of missing data occurs for conceptual rather than measurement error reasons – data were not available for two subjects for ‘On activity’, ‘At rest’ and the mean, but this was for logistical and organisational reasons rather than as threats to the conceptual relevance of the variable. It was therefore not appropriate to discard a variable on this basis.

b) where the incidence of the variable in the cohort was less than 5% – this was not applicable

c) where multiple variables can be meaningfully collapsed down into a single score – this was applicable

d) where there is high correlation between variables within a category indicating that a single measure would be representative – this was applicable

Combination of c) and d):

As a continuous variable, pain as measured by VAS was found to be non-normally distributed (see appendix 6; although mean of the three scores was normally distributed). Therefore Spearman's rho was used to statistically assess for any relationship between the variables (Altman 1991).

Table 6-35: Correlation between VAS variables

VAS variable	On activity	At rest	At night	Mean
On activity	1	0.346 *	0.640 *	0.821 *
At rest	0.346 *	1	0.475 *	0.678 *
At night	0.640 *	0.475 *	1	0.891 *
Mean	0.821 *	0.678 *	0.891 *	1

Key: Correlation coefficient r, * = significant correlation at $p<0.01$

Using Spearman's rho a statistically significant relationship between all variables was demonstrated and the correlation coefficients were all positive.

The three highest correlations all involved the mean of the three scores and so this was selected to be representative of the category. From a conceptual perspective this was advantageous because it encompasses all three aspects of pain presentation relative to the patient.

6.6.4 C2. Patient reported measures: Psychological symptoms

1. Clinical relevance – the gradings for each patient under the individual categories of Distress, Depression, Anxiety, and Somatisation all have inherent clinical relevance

2. Conceptual relevance:

a) where a high % of missing data occurs for conceptual rather than measurement error reasons – data were not available for two subjects but this was for logistical and organisational reasons rather than as threats to the conceptual relevance of the variable. It was therefore not appropriate to discard a variable on this basis.

b) where the incidence of the variable in the cohort was less than 5% – incidence of 'strongly elevated' for anxiety and depression; therefore a mechanism was sought to retain this data.

c) where multiple variables can be meaningfully collapsed down into a single score – this was applicable

d) where there is a high correlation between variables within a category indicating that a single measure would be representative – this was applicable

Combination of b), c) and d):

To provide a quantitative estimate of any correlation between variables, Spearman's rho was run. However as Spearman's rho can struggle where there are a number of ties in the data, Kendall's tau was run in parallel (Altman 1991).

Table 6-36: Correlation between 4DSQ variables (Spearman's rho)

4DSQ variable	Distress	Depression	Anxiety	Somatisation
Distress	1	0.688 *	0.555 *	0.475 *
Depression	0.688 *	1	0.569 *	0.507 *
Anxiety	0.555 *	0.569 *	1	0.417 *
Somatisation	0.475 *	0.507 *	0.417 *	1

Key: Correlation coefficient r, * = significant correlation at p<0.001

Table 6-37: Correlation between 4DSQ variables (Kendall's tau)

4DSQ variable	Distress	Depression	Anxiety	Somatisation
Distress	1	0.673 *	0.544 *	0.461 *
Depression	0.673 *	1	0.556 *	0.496 *
Anxiety	0.544 *	0.556 *	1	0.409 *
Somatisation	0.461 *	0.496 *	0.409 *	1

Key: Correlation coefficient r, * = significant correlation at p<0.001

Both Spearman's rho and Kendal's Tau demonstrated statistically significant, consistent relationships between each of the categories; and the strength of these was largely comparable.

The collapsing of categories into a single score has been previously used in the shoulder literature to present psychological symptom prevalence (Koorevaar et al. 2016). It was therefore deemed appropriate to collapse the 4DSQ data for the cohort down into a single binary score such that where no categories had an elevated reading the patient was coded as 0, but where patients had one or more category with an elevated reading they were coded as 1 for psychological symptoms. This data can be seen in appendix 6 '4DSQ: dichotomised score', whereby 49 subjects (64.5%) had no

categories with an elevated value, 25 subjects (32.9%) had 1 or more category with an elevated reading and 2 subjects (2.6%) had no 4DSQ data.

6.6.5 C3. Patient reported measures: Function / Disability

1. Clinical relevance – the total SPADI score and the two values (total pain score and total disability score) that comprise it all have inherent clinical relevance.

2. Conceptual relevance:

a) where a high % of missing data occurs for conceptual rather than measurement error reasons – this was not applicable.

b) where the incidence of the variable in the cohort was less than 5% – this was not applicable.

c) where multiple variables can be meaningfully collapsed down into a single score – whilst the total SPADI score is commonly reported, the component parts of the total score are also commonly reported, including in the shoulder prognostic study by Kromer et al. (2014). It was therefore not appropriate to discard a variable solely on this basis.

d) where there is high correlation between variables within a category indicating that a single measure would be representative – see below

As continuous variables, total pain score, total disability score and total SPADI score were all found to be normally distributed (see appendix 6) and so Pearson's correlation coefficient was therefore used to statistically assess for any relationship between the variables (Altman 1991).

Table 6-38: Correlation between SPADI variables

SPADI variable	Total pain	Total disability	Total SPADI
Total pain	1	0.800 *	0.919 *
Total disability	0.800 *	1	0.970 *
Total SPADI	0.919 *	0.970 *	1

Key: Correlation coefficient r, * = significant correlation at $p<0.001$

Pearson's correlation coefficient demonstrated statistically significant ($p<0.001$) relationships with high correlation coefficients between each of the variables. The high correlations between total SPADI score and the total pain and total disability scores are to be expected as each are the component parts of the SPADI score. The high correlation between the component scores and the total SPADI score indicate that any single measure could be representative of the category. As it incorporates both pain and disability, the total SPADI score was selected.

6.6.6 D1. Clinical measures: Strength

1. Clinical relevance – the mean moment data for IR, ER and Scaption on the asymptomatic side were discarded as the asymptomatic measures were deemed of limited relevance to the presenting condition and were therefore unlikely to provide meaningful prognostic information.

2. Conceptual relevance:

a) where a high % of missing data occurs for conceptual rather than measurement error reasons – Missing data were present for subject M2279h across all strength measures but this was due to technical issues. The only situation where a reading was not possible was subject F4417h who was unable to reach the unit's recording threshold due to pain during the test movement. Although this is a conceptual issue in relation to the ability for a cohort patient to have a meaningful data point, the low incidence means that it was not deemed appropriate to exclude any variables in the strength category due to this criterion.

b) where the incidence of the variable in the cohort was less than 5% – this was not applicable.

c) where multiple variables can be meaningfully collapsed down into a single score – this was not applicable.

d) where there is high correlation between variables within a category indicating that a single measure would be representative – see below

As continuous variables, all strength variables were found to be non-normally distributed (see appendix 6) and so Spearman's rho was therefore used to statistically assess for any relationship between the variables (Altman 1991).

Table 6-39: Correlation between strength variables

Strength variable	IR mean moment	ER mean moment	Scap mean moment
IR mean moment	1	0.899 *	0.772 *
ER mean moment	0.899 *	1	0.811 *
Scap mean moment	0.772 *	0.811 *	1

Key: Correlation coefficient r, * = significant correlation at p<0.001, Scap = scapular

Pearson's correlation coefficient demonstrated statistically significant (p<0.001) relationships with high correlation coefficients between each of the variables. The highly positive correlations across all variables indicated that a single measure could be representative of the category. ER was

therefore selected due to its clinical relevance as a representation of the posterior rotator cuff function (Reinold et al. 2004) and due to the common prescribing of resisted external rotation strengthening exercises in SIS/RCTendinopathy non-invasive, multimodal physiotherapy (Holmgren et al. 2012).

6.6.7 D2. Clinical measures: ROM

1. Clinical relevance – the limit of ROM data for the asymptomatic side were discarded as the asymptomatic measures were deemed of limited relevance to the presenting condition and were therefore unlikely to provide meaningful prognostic information.

2. Conceptual relevance:

a) where a high % of missing data occurs for conceptual rather than measurement error reasons – Missing data were present for 10 subjects for the variable of 'point of symptom exacerbation by movement'. The reasons for this were either that the subject experienced no symptoms during movement (i.e. ceiling effect) or that their symptoms remained the same throughout the movement (i.e. floor effect). These were both clinically meaningful threats to the conceptual relevance of the variable of point of symptom exacerbation by movement.
b) where the incidence of the variable in the cohort was less than 5% – this was not applicable.
c) where multiple variables can be meaningfully collapsed down into a single score – this was not applicable.
d) where there is a high correlation between variables within a category indicating that a single measure would be representative – recognising the potential threat posed by the inclusion of the variable of 'point of symptom exacerbation by movement', potential association with the variable 'limit of ROM' was explored.

As a continuous variable, limit of ROM was not normally distributed (see appendix 6) and so Spearman's rho was used to statistically assess for any relationship between the variables (Altman 1991).

Table 6-40: Correlation between ROM variables

ROM variable	Symptomatic side: limit of ROM	Symptomatic side: point of symptom exacerbation by movement
Symptomatic side: limit of ROM	1	0.611 *
Symptomatic side: point of symptom exacerbation by movement	0.611 *	1

Key: Correlation coefficient r, * = significant correlation at p<0.001

Spearman's rho demonstrated a statistically significant (p<0.001) relationship between the two variables with a correlation coefficients of 0.611. The positive correlation indicated that either measure could be representative of the category. Due to the conceptual issues identified with the variable 'point of symptom exacerbation by movement', the variable 'limit of ROM' was selected as the candidate variable for the category of strength.

6.6.8 D3. Clinical measures: Scapular movement and control

1. Clinical relevance – the gradings from both the un-loaded and loaded trials have inherent clinical relevance and so it was therefore not appropriate to discard a variable on this basis.

2. Conceptual relevance:

a) where a high % of missing data occurs for conceptual rather than measurement error reasons – 42% of the cohort at baseline had no data for the loaded trial, all because of symptom levels or exacerbation from the non-loaded trial. This prevented them from performing the loaded trial. Such data is evidence of a threat to the conceptual relevance of using the loaded trial as a potential prognostic factor.
b) where the incidence of the variable in the cohort was less than 5% – this was not applicable.
c) where multiple variables can be meaningfully collapsed down into a single score – this was not applicable.
d) where there is high correlation between variables within a category indicating that a single measure would be representative – see below.

To provide a quantitative estimate of any correlation between variables, Spearman's rho was run. However as Spearman's rho can become inaccurate where there are a number of ties in the data, Kendall's tau was run in parallel (Altman 1991).

Table 6-41: Correlation between scapular variables (Spearman's rho)

Scapular variable	Un-loaded SDT	Loaded SDT
Un-loaded SDT	1	0.257 *
Loaded SDT	0.257 *	1

Key: Correlation coefficient r, * = significant correlation at p<0.05

Table 6-42: Correlation between scapular variables (Kendall's tau)

Scapular variable	Un-loaded SDT	Loaded SDT
Un-loaded SDT	1	0.259 *
Loaded SDT	0.259 *	1

Key: Correlation coefficient r, * = significant correlation at p<0.05

Both Spearman's rho and Kendal's Tau demonstrated statistically significant relationships in the same direction between each of the categories and that the strength of these was largely comparable.

Due to the statistically significant relationship between the un-loaded and loaded data and the large percentage of missing data from the loaded variable due to conceptual reasons, it was deemed appropriate to use the un-loaded SDT data as the representative variable for the category of scapular control and movement.

6.6.9 E1. Structural pathology via imaging

1. Clinical relevance – Whilst acknowledging the non-definitive link between structural pathology and symptoms, each of the sonographic diagnosis categories have inherent potential clinical relevance. It was therefore not appropriate to discard a variable on this basis.

2. Conceptual relevance:

a) where a high % of missing data occurs for conceptual rather than measurement error reasons – this was not applicable.

b) where the incidence of the variable in the cohort was less than 5% – The sonographic diagnosis categories of PTT and FTT had an incidence in the main cohort of 2.6% and 3.9% (respectively). Applying the above data reduction rule would mean that such findings would be discarded.

However from a conceptual and clinical perspective this would be questionable because from the perspective of the conceptual framework around structural pathology such sonographic findings could be considered highly relevant. To address this, the following data reduction rule was applied:

c) where multiple variables can be meaningfully collapsed down into a single score – in order to retain the low frequency variables and partially address concerns regarding the reproducibility of the tendinopathy data, conversion of the imaging data into a composite, ordinal variable was performed. This was 0 = no pathology, 1 = one of bursal or cuff pathology, 2 = both bursal and cuff pathology.

This new variable was generated from the existing imaging data and can be seen in the last column of the sonographic diagnosis data in appendix 6. The frequency of these categories was as follows: No grading possible (n=1; 1.3%), no pathology (n=13; 17.1%), one of bursal or cuff pathology (n=38; 50%), both bursal and cuff pathology (n=24; 31.6%).

6.6.10 E1. Structural pathology via orthopaedic tests

1. Clinical relevance – The individual findings for Neer's sign, the Empty can and Hawkins-Kennedy test and pain on active abduction have inherent clinical relevance. However a battery of tests approach is commonly advocated as a mechanism to address the sensitivity/specificity issues of the individual tests (Park et al. 2005; Hegedus et al. 2012; Diercks et al. 2014).

2. Conceptual relevance:

a) where a high % of missing data occurs for conceptual rather than measurement error reasons – no further data reduction applicable.

b) where the incidence of the variable in the cohort was less than 5% – this was not applicable

c) where multiple variables can be meaningfully collapsed down into a single score – acknowledging the battery of tests approach, a cut off point of 3 or more positive tests (Michener et al. 2009) was chosen. This provided a meaningful mechanism for reducing the number of potential candidate variables from 4 binary variables (Neer's sign, the Empty can and Hawkins-Kennedy test and pain on active abduction) to 1 binary variable (yes/no: 3 or more +ve test results).

This new variable was generated from the existing orthopaedic test data and can be seen in the last column of the orthopaedic test data in appendix 6. The frequency of this category was three or more +ve test results: Yes (n=56; 73.7%), No (n=20; 26.3%), demonstrating that just under 3/4 of subjects had three or more positive findings.

6.7 Tapered data reduction aligned with sample size

As previously noted, 76 patients were eligible and consented to the study, which was less than the sample size of 100. Due to the emphasis on 10 cases per variable (Peduzzi et al. 1996) only seven variables were permitted to be entered into the prognostic modelling stage. Using the same principles, further variable trimming was undertaken, including the retention of a minimum of one variable per category.

6.7.1 Patient reported and clinical measures categories

In the patient reported measures category there was the potential for a conceptual overlap between pain (as measured by mean VAS) and function / disability (as measured by total SPADI score). In particular the fact that 5/13 of the total SPADI score is derived from questions specific to pain. As mean of the three VAS scores and the total SPADI score were both normally distributed, Pearson's correlation coefficient was used to statistically assess for any relationship between the variables (Altman 1991).

Table 6-43: Correlation between patient reported variables

Patient reported variable	Mean of the 3 VAS scores	Total SPADI score
Mean of the 3 VAS scores	1	0.800 *
Total SPADI score	0.800 *	1

Key: Correlation coefficient r, * = significant correlation at $p<0.001$

The results show a statistically significant, strong and positive correlation between VAS Mean and SPADI Mean. As total SPADI score also measures elements of disability then it was selected over mean of the three VAS scores because total SPADI score is more likely to provide broadly representative information due to the disability components of the outcome measure.

Between the patient reported and clinical measures categories there was the potential for overlap between the total SPADI score and both strength and ROM. This is because 8/13 of the total SPADI score is derived from questions specific to physical function, which is in part a composite of shoulder strength and ROM. As neither ER mean moment nor limit of ROM were normally distributed, any correlation between them was assessed via Spearman's rho (Altman 1991).

Table 6-44: Correlation between patient reported and clinical measure variables

Patient reported variable	ER mean moment	Total SPADI score
ER mean moment	1	-0.259 *
Total SPADI score	-0.259 *	1

Key: Correlation coefficient r, * = significant correlation at p<0.05

Patient reported variable	Limit of ROM	Total SPADI score
Limit of ROM	1	-0.505 **
Total SPADI score	-0.505 **	1

Key: Correlation coefficient r, ** = significant correlation at p<0.001

The results show that both ER mean moment and limit of ROM had negative correlations with SPADI mean. The inverse relationships are explained by a higher SPADI (i.e. more disability and pain) being associated with less ROM and strength (i.e. a lower ROM and strength value). The stronger and more statistically significant association between limit of ROM and total SPADI score indicated the greater correlation. Therefore limit of ROM was selected as the variable to remove.

6.7.2 Structural pathology category

In the structural pathology category the decision to retain either Ultrasound or Orthopaedic tests was informed by evidence in the published literature of correlation between ultrasound findings and orthopaedic tests (Naredo et al. 2002; Micheroli et al, 2015) including superior sensitivity and specificity of the former over the later (Iagnocco et al. 2003; Kelly et al. 2010). This evidence combined with the novel inclusion of ultrasound as a potential prognostic factor in this area of research informed the decision to retain the ultrasound findings in the prognostic model but not the orthopaedic test findings.

6.8 Variables to be entered into the prognostic modelling stage

The seven variables were therefore:

A. Demographics = Age (continuous variable)

B. Clinical history = Symptom duration (ordinal variable)

C. Patient reported measures = Psychological symptoms via 4DSQ (dichotomised variable); Function / Disability via total SPADI score (continuous variable)

D. Clinical measures = Strength via symptomatic side ER mean moment (continuous variable); Scapular movement and control via unloaded SDT (ordinal variable)

E. Structural pathology = Ultrasound evidence of pathology (ordinal variable)

Noting the conceptual frameworks presented in the introduction chapter, these map as follows:

- Patho-anatomical model:
    - Intrinsic factors = E. Structural pathology
    - Extrinsic factors = D. Clinical measures; specifically scapular movement and control
- Psycho-social model = C. Patient reported measures; specifically 4DSQ
- ICF classification:

  Functioning and disability:
    - Body function: Mental function = C. Patient reported measures; specifically 4DSQ
    - Body function: Sensory function and Pain = C. Patient reported measures; specifically pain (represented by total SPADI score)
    - Body function: Neuromusculoskeletal and movement related functions = D. Clinical measures; specifically strength, ROM and scapular movement and control (represented by strength and scapular movement and control)
    - Body structure: Structure related to movement = E. Structural pathology
    - Activities and participation = C. Patient reported measures; specifically SPADI

  Contextual factors:
    - Personal factors = A. Demographics; specifically age

  Qualifier:
    - Duration = B. Clinical history; specifically duration

6.9 Differences between those lost to follow up and those where follow up data was available

Having identified the seven variables to be entered into the prognostic modelling stage, the potential that those lost to follow up may have skewed the subsequent prognostic models was explored. Due to the small numbers lost at discharge (n=3: M2028w, F2197w and M9819h) and 3 months post-discharge (n=11: M2146h, F1634w, F9939h, M5703h, F2387h, F0732w, M6858w, F1188h, F1669h, F5690h, M3609h) it was deemed appropriate to use only descriptive statistics to explore any systematic bias for the seven variables and absolute baseline OSS.

Below are the descriptive statistics (relative to the data type and distribution) for each of the seven variables and absolute baseline OSS. Appendix 6 contains the variable values for the patients lost to follow at each of the two time points.

6.9.1 Baseline OSS

Table 6-45: Non-normally distributed; continuous data (OSS)

	Median	25th percentile	75th percentile
Available for follow up at discharge (n=73)	33	26	37
Available for follow up at 3 months post-discharge (n=62)	33	26	37

As per appendix 6, baseline OSS was found to be non-normally distributed. For those patients not available at discharge (n=3; appendix 6), one was at the 75th percentile (M2028w = OSS of 37) and one was above the 75th percentile (F2197w = OSS of 42). Therefore those recruited to the study but unavailable for the baseline to discharge prognostic model had baseline function (as measured by the OSS) at the upper end of those who were available for the baseline to discharge prognostic model. However the likelihood of this skewing the data is limited by the small sample number who were elevated (n=2).

For those patients not available at 3 months post-discharge (n=14, i.e. those lost at discharge *plus* those lost at 3 months post-discharge; appendix 6), four were above the 75th percentile (OSS of 38, 39, 40 and 42) but two were well below the 25th percentile (OSS of 19 and 21). Therefore the likelihood that baseline function as measured by the OSS for those lost to follow up skewed the data is small.

6.9.2 Age

Table 6-46: Non-normally distributed; continuous data (Age / years)

	Median	25th percentile	75th percentile
Available for follow up at discharge (n=73)	50	43	62
Available for follow up at 3 months post-discharge (n=62)	52	47	62

As per appendix 6, age was found to be non-normally distributed. For those patients not available at discharge, two were below the 25th percentile (27 and 29 years old) whilst the other subject was

above the median age but within the 75th percentile. Again, with the small sample size the likelihood of skewing of the data due to age distribution is small.

For those patients not available at 3 months post-discharge, 1 (66 years) was above the 75th percentile but 8 (18, 19, 27, 29, 30, 34, 38 and 40 years) were below the 25th percentile. Therefore, there is the potential for a systematic bias whereby a disproportionate number of those unavailable for the 3 months post-discharge model were younger patients.

6.9.3 Symptom duration

Table 6-47: Ordinal data (Duration of symptoms)

	0 to <3 months	3 to <6 months	6 to <12 months	12 to <24 months	24 months or longer
Available for follow up at discharge (n=73)	5 (6.8%)	23 (31.5%)	26 (35.6%)	7 (9.6%)	12 (16.4%)
Available for follow up at 3 months post-discharge (n=62)	5 (8.1%)	22 (35.5%)	20 (32.3%)	6 (9.7%)	9 (14.5%)

For those patients not available at discharge, they were spread between 0 to <3 months, 3 to <6 months and 12 to <24 months. The broad spread and small sample size make it unlikely that any skewing of the sample was present due to the duration of symptoms of those not available for the baseline to discharge model.

For those patients not available at 3 months post-discharge, they were spread between 0 to <3 months (n=1), 3 to <6 months (n=2), 6 to <12 months (n=6), 12 to <24 months (n=2) and 24 months or longer (n=3). The broad spread across the categories for this variable makes it unlikely that any skewing of the sample occurred due to the duration of symptoms of those not available for the baseline to 3 months post-discharge model.

6.9.4 Psychological symptoms via 4DSQ

Table 6-48: Dichotomised data (4DSQ)

	No categories with elevated psychological symptoms	1 or more category with elevated psychological symptoms
Available for follow up at discharge (n=71)	46 (64.8%)	25 (35.2%)
Available for follow up at 3 months post-discharge (n=61)	41 (67.2%)	20 (32.8%)

For those patients not available at discharge (n=3), all of them were classed as having no categories with elevated psychological symptoms. Almost 2/3 of subjects available for follow up at discharge were also classed as such and so along with the small sample size this is unlikely to represent a systematic bias.

For those patients not available at 3 months post-discharge, n=8/13 (61.5%) were classed as having no categories with elevated psychological symptoms which is very comparable to those available for follow up at 3 months post-discharge (67.2%). As such, skewing of the sample due to the psychological symptoms of those not available for the baseline to 3 months post-discharge model is unlikely.

6.9.5 Total SPADI score (%)

Table 6-49: Normally distributed; continuous data (Total SPADI / %)

	Mean	95% CI (lower bound)	95% CI (upper bound)
Available for follow up at discharge (n=73)	43.5	38.5	48.6
Available for follow up at 3 months post-discharge (n=62)	44.4	38.6	50.2

Key: 95% CI = 95% confidence interval of the mean

As per appendix 6, total SPADI was found to be normally distributed. For those patients not available at discharge, one had a total SPADI score (54.6%) that was above the upper 95% CI and 1 had a score

(8.5%) much lower than the lower 95% CI. Acknowledging the small sample size then this is unlikely to represent a systematic bias.

For those patients not available at 3 months post-discharge, 3 (54.6%, 54.6% and 60.0%) were above the upper 95% CI and 6 (8.5%, 14.6%, 16.9%, 27.7%, 35.4%, 36.2%) had a score below the lower 95% CI. Along with those patients not available at 3 months post-discharge having a mean of 38.1% this provides limited evidence of a systematic bias whereby a disproportionate proportion of those unavailable for the 3 months post-discharge model had lower total SPADI scores and so had lower pain and disability scores at baseline.

6.9.6 Symptomatic side ER mean moment (Nm)

Table 6-50: Non-normally distributed; continuous data (Mean moment / Nm)

	Median	25th percentile	75th percentile
Available for follow up at discharge (n=73)	8.9	7.0	12.6
Available for follow up at 3 months post-discharge (n=62)	8.9	7.1	12.7

As per appendix 6, Symptomatic side ER mean moment was found to be non-normally distributed. For those patients not available at discharge, one was below the 25th percentile (2.2Nm) whilst the other subjects were above the 75th percentile (13.0Nm and 18.8Nm). With the small sample size, the likelihood of skewing of the data due to strength distribution is small.

For those patients not available at 3 months post-discharge, five were above the 75th percentile and five were below the 25th percentile. Although this shows a wide spread of data in those not available at 3 months post-discharge, a directional bias is unlikely.

6.9.7 Scapular movement and control via unloaded SDT

Table 6-51: Ordinal data (SDT)

	Normal	Subtle abnormality	Obvious abnormality
Available for follow up at discharge (n=73)	26 (35.6%)	28 (38.4%)	19 (26.0%)
Available for follow up at 3 months post-discharge (n=62)	21 (33.9%)	25 (40.3%)	16 (25.8%)

For those patients not available at discharge, all were graded as having subtle abnormalities. Along with the small sample size, the mid-grading of all three subjects makes it unlikely that any skewing of the sample due to scapular movement and control of those not available for the baseline to discharge model.

For those patients not available at 3 months post-discharge they were spread between categories of normal (n=5), subtle abnormalities (n=6) and obvious abnormalities (n=3). These largely reflect the distribution of those available for follow up at 3 months post-discharge, making it unlikely that any skewing of the sample due to the scapular movement and control of those not available for the baseline to 3 months post-discharge model.

6.9.8 Ultrasound evidence of pathology

Table 6-52: Ordinal data (U/S)

	No pathology	1 of bursal or cuff pathology	Both bursal and cuff pathology
Available for follow up at discharge (n=72)	12 (16.7%)	36 (50%)	24 (33.3%)
Available for follow up at 3 months post-discharge (n=61)	11 (18.0%)	29 (47.5%)	21 (34.5%)

For those patients not available at discharge, one had 1 of bursal or cuff pathology and two had no pathology. Although this represents slight skewing towards less evidence of pathology on ultrasound for those not available compared to those available, the small sample size makes it unlikely to represent any systematic bias.

For those patients not available at 3 months post-discharge they were spread between categories of no pathology (n=4), 1 of bursal or cuff pathology (n=6) and both bursal and cuff pathology (n=4). These largely reflect the distribution of those available for follow up at 3 months post-discharge, making it unlikely that any skewing of the sample due to the evidence of pathology on ultrasound of those not available for the baseline to 3 months post-discharge model.

6.10 Regression analysis

The prognostic data analysis approach detailed in the method section was applied to generate two separate models: one for baseline to discharge and one for baseline to 3 months post-discharge.

6.10.1 Regression analysis for baseline to discharge

6.10.1.1 Components of the regression model for OSS change score from baseline to discharge

Using the criteria and method previously stipulated, the model where the dependent variable was the OSS change score from baseline to discharge comprised total SPADI as block 1 and total SPADI combined with Age as block 2:

Table 6-53: significant regression analysis findings for baseline to discharge change in OSS

Variable at baseline	R	R^2	R^2 adjusted	F-ratio	F significance	Durbin-Watson
Total SPADI	0.301	0.090	0.077	6.653	0.012	2.228
Total SPADI & Age	0.397	0.157	0.132	6.166	0.004	

Variable at baseline	β	β significance	95% CI: Lower bound	95% CI: Upper bound	Collinearity statistic: Tolerance	Collinearity statistic: VIF
Total SPADI	0.301	0.012	0.026	0.203	1.000	1.000
Total SPADI & Age	0.367	0.002	0.051	0.228	0.938	1.066
	-0.268	0.025	-0.310	-0.021	0.938	1.066

Key: 95% CI = 95% confidence intervals, VIF = Variance inflation factor

6.10.1.2 Explanation of prognostic model parameters

The R^2 value for block 1 (total SPADI) was 0.090 and for block 2 was 0.157. This indicates that total SPADI at baseline accounted for 9% of the variation in OSS change. When combined with Age, they

collectively accounted for 15.7% of the variance in OSS change. Therefore entering Age into the model in block 2 (total SPADI *and* Age) accounted for an extra 6.7% of the variance in the OSS scores. The adjusted R^2 value for block 2 is 0.157 – 0.132 = 0.025 or 2.5% and indicates that deriving the model from the population rather than a sample would cause it to account for approx. 2.5% less variance in the outcome.

The F-ratio from the ANOVA was statistically significant for both block 1 (p=0.012) and block 2 (p=0.004) at 6.653 and 6.166; respectively. Both F-ratios are greater than 1 demonstrating that there is a significant improvement in the final model predicting outcome, calculated by the ratio of the improvement in predicting as a consequence of the model to the residual inaccuracy in the model.

The parameters of the model demonstrated via the coefficients (β) for each predictor that for total SPADI it is +ve and Age it is –ve. Consequently, there is a positive relationship between total SPADI and change in OSS score meaning that a higher total SPADI (greater pain and disability) was associated with a greater improvement in OSS. Conversely, the negative relationship between Age and change in OSS score means that greater age was associated with less improvement in OSS. The significance values of the coefficients (β) were p=0.002 and p=0.025; respectively. These confirm that total SPADI contributed more to the model than Age. The 95% CI's for each variable are relatively narrow and do not cross zero, supporting the assumption that the estimates from the model are likely to approximate to the true population values (Field 2009).

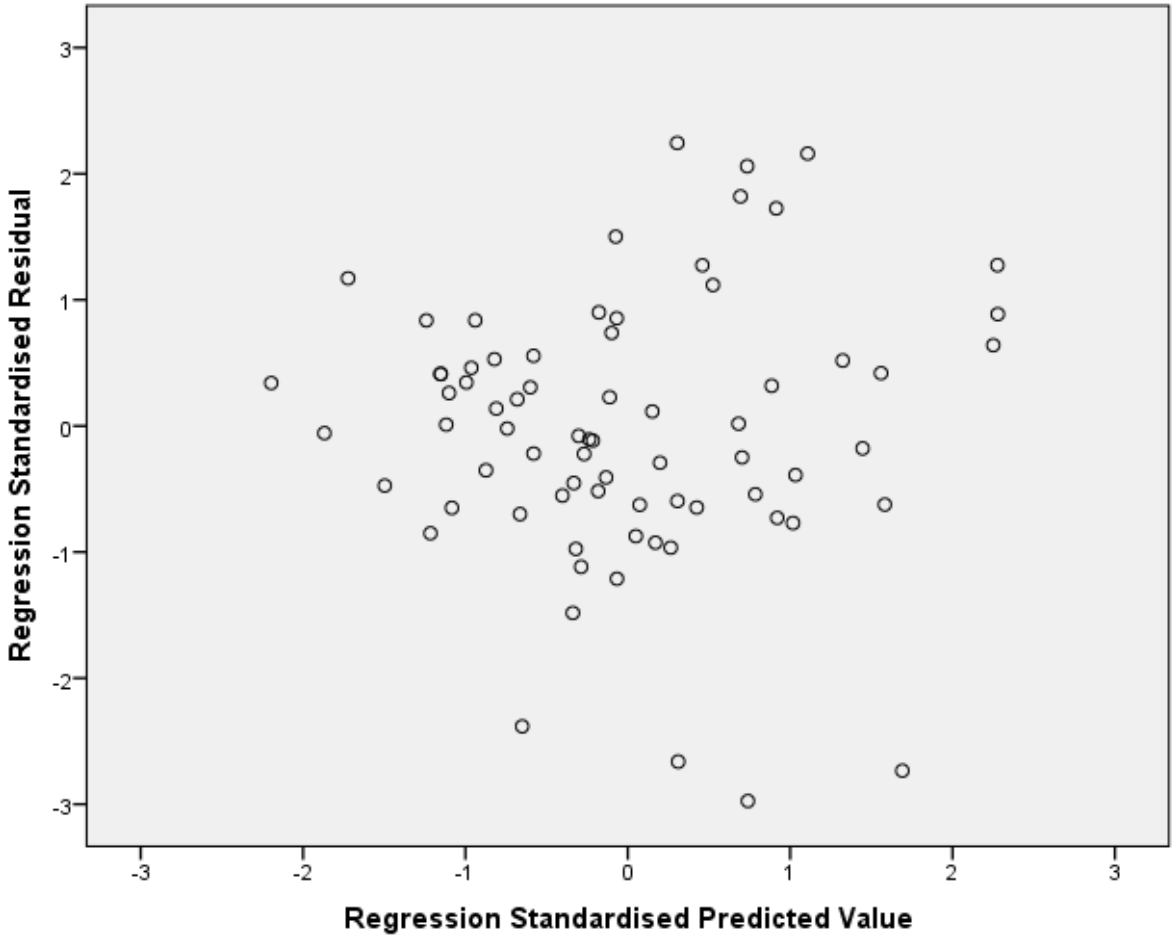
Where total SPADI was entered into the model (i.e. block 1) each of the variables not included in the first stage of the model have p values of greater than 0.05. Specifically these were symptom duration p=0.692, psychological symptoms via 4DSQ p=0.709, Symptomatic side ER mean moment p=0.961, Scapular movement and control via unloaded SDT p=0.275 and Ultrasound evidence of pathology p=0.192. Conversely, Age was significant with p=0.025 and so was entered into the model in the next block.

In the final version of the model (i.e. block 2) the excluded variables had p values greater than 0.05. Specifically symptom duration p=0.887, psychological symptoms via 4DSQ p=0.267, Symptomatic side ER mean moment p=0.896, Scapular movement and control via unloaded SDT p=0.462 and Ultrasound evidence of pathology p=0.403.

6.10.1.3 Tests of the fit of the model and diagnostics

The proximity of the Durbin-Watson value to 2 indicates that the residuals are not correlated (Field 2009).

Figure 6-3: Graph of the standardised residuals (errors) against the standardised predicted values



The random distribution of the data in the above graph, with an absence of curve indicates that the assumptions of linearity and homoscedasticity have been met (Field 2009).

The collinearity statistics (VIF <10; average VIF close to 1, tolerance above 0.2) indicate that there is no collinearity in the data (Meyers 2000).

Figure 6-4: Histogram of the residuals

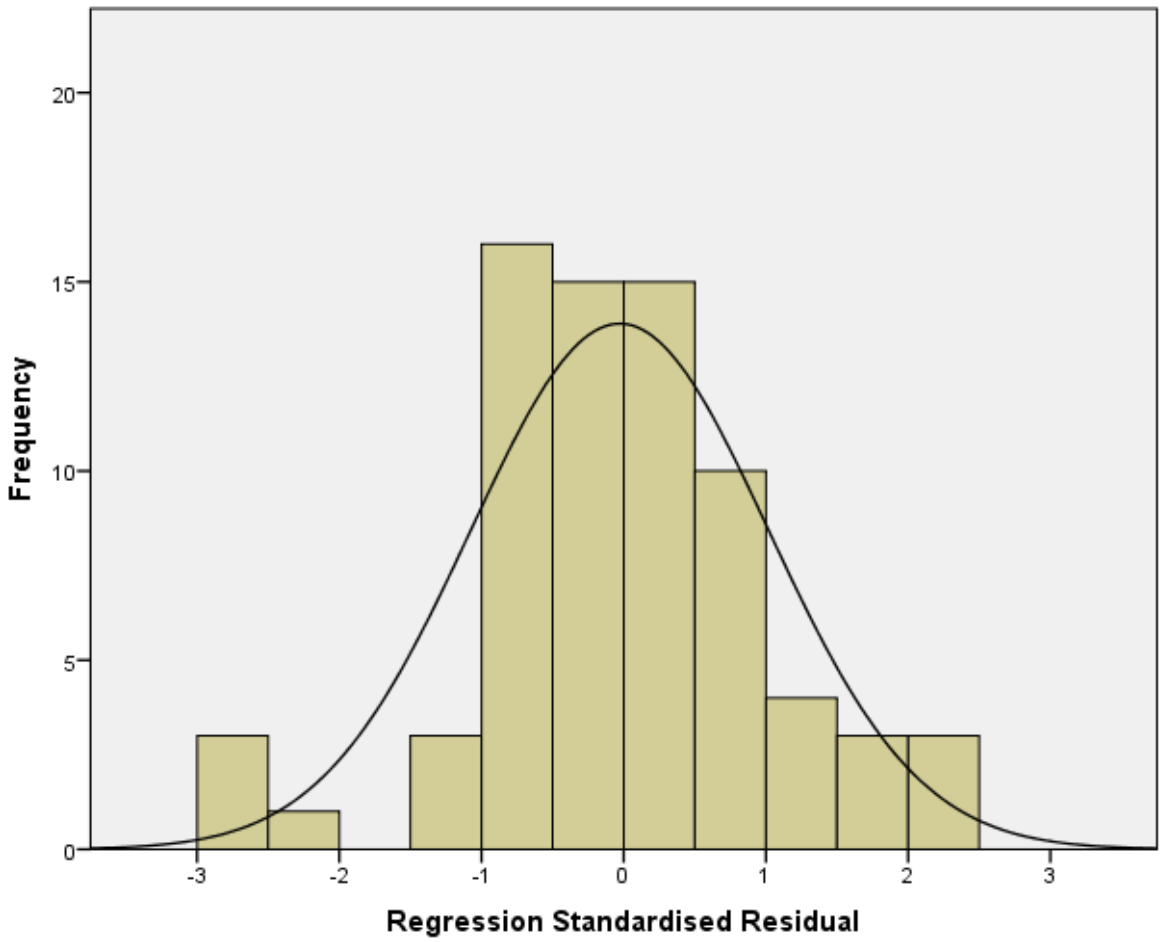
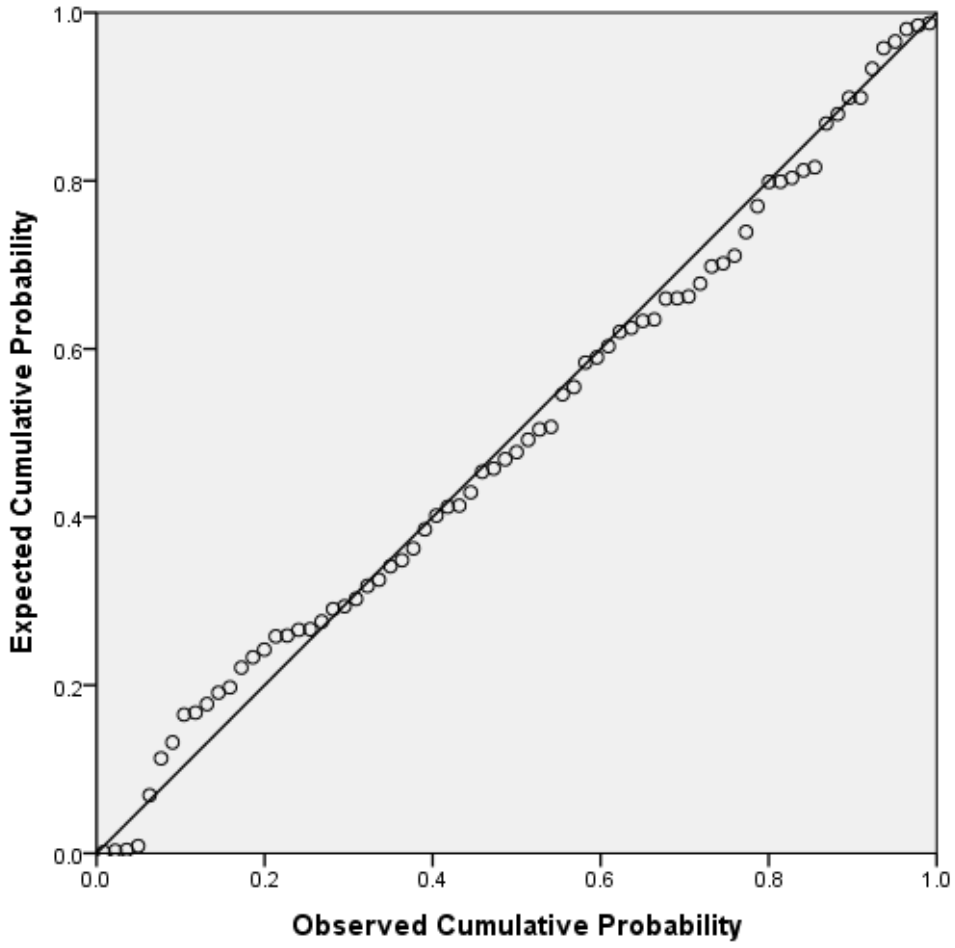


Figure 6-5: PP plot of the residuals



The close approximation of the histogram plot to a normal distribution, along with the close approximation of the PP plot to the reference line confirm the approximately normal distribution of the residuals.

6.10.2 Regression analysis for baseline to 3 months post-discharge

6.10.2.1 Components of the regression model for OSS change score from baseline to 3 months post-discharge

Using the criteria and method previously stipulated, only total SPADI was retained in the model where the dependent variable was the OSS change score from baseline to 3 months post-discharge.

Table 6-54: significant regression analysis findings for baseline to 3 months post-discharge change in OSS

Variable at baseline	R	R^2	R^2 adjusted	F-ratio	F significance	Durbin-Watson
SPADI mean	0.310	0.096	0.081	6.081	0.017	2.607

Variable at baseline	β	β significance	95% CI: Lower bound	95% CI: Upper bound	Collinearity statistic: Tolerance	Collinearity statistic: VIF
SPADI mean	0.310	0.017	0.020	0.196	1.000	1.000

Key: 95% CI = 95% confidence intervals, VIF = Variance inflation factor

6.10.2.2 Explanation of prognostic model parameters

The R^2 value for the model was 0.096, which indicates that total SPADI at baseline accounts for 9.6% of the variation in OSS change. The adjusted R^2 value for the model is 0.096 – 0.081 = 0.015 or 1.5% and indicates that deriving the model from the population rather than a sample would cause it to account for approximately 1.5% less variance in the outcome.

The F-ratio from the ANOVA was statistically significant for the model (p=0.017) at 6.081. As the F-ratio is greater than 1 then there is a significant improvement in the final model predicting outcome, calculated by the ratio of the improvement in predicting as a consequence of the model to the residual inaccuracy in the model.

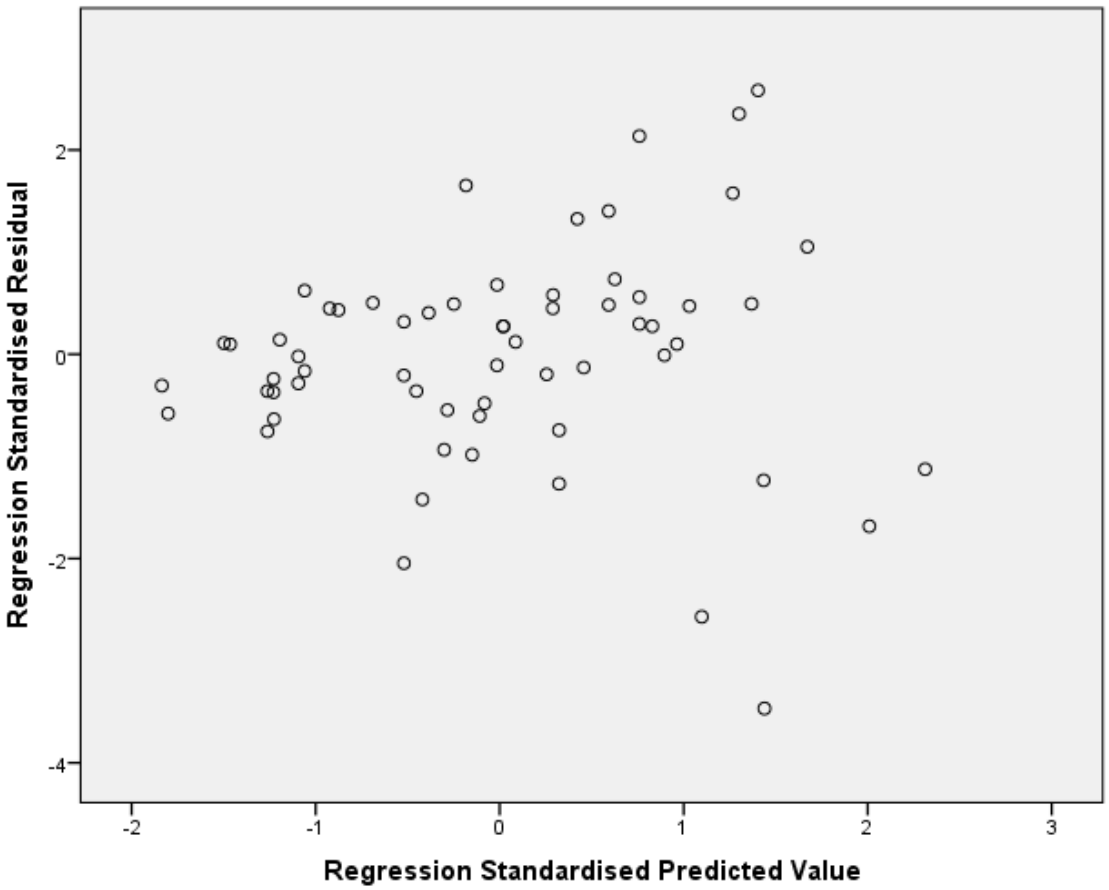
The parameter of the model demonstrates via the coefficient (β) for the predictor that for total SPADI it is +ve. Consequently there is a positive relationship between total SPADI and change in OSS score meaning that a higher total SPADI (greater pain and disability) was associated with a greater improvement in OSS. The 95% CI's for the variable are relatively narrow and do not cross zero, supporting the assumption that the estimates from the model are likely to approximate to the true population values (Field 2009).

Where total SPADI was entered into the model, each of the variables not included had p values of greater than 0.05. Specifically these were Age p=0.199, symptom duration p=0.483, psychological symptoms via 4DSQ p=0.682, Symptomatic side ER mean moment p=0.733, Scapular movement and control via unloaded SDT p=0.156 and Ultrasound evidence of pathology p=0.186.

6.10.2.3 Tests of the fit of the model and diagnostics

The proximity of the Durbin-Watson value to 2 indicates that the residuals are not correlated (Field 2009).

Figure 6-6: Graph of the standardised residuals (errors) against the standardised predicted values



The random distribution of the data in the above graph, with an absence of curve indicates that the assumptions of linearity and homoscedasticity have been met (Field 2009).

The collinearity statistics (VIF <10; average VIF close to 1, tolerance above 0.2) indicate that there is no collinearity in the data (Meyers 2000).

Figure 6-7: Histogram of the residuals

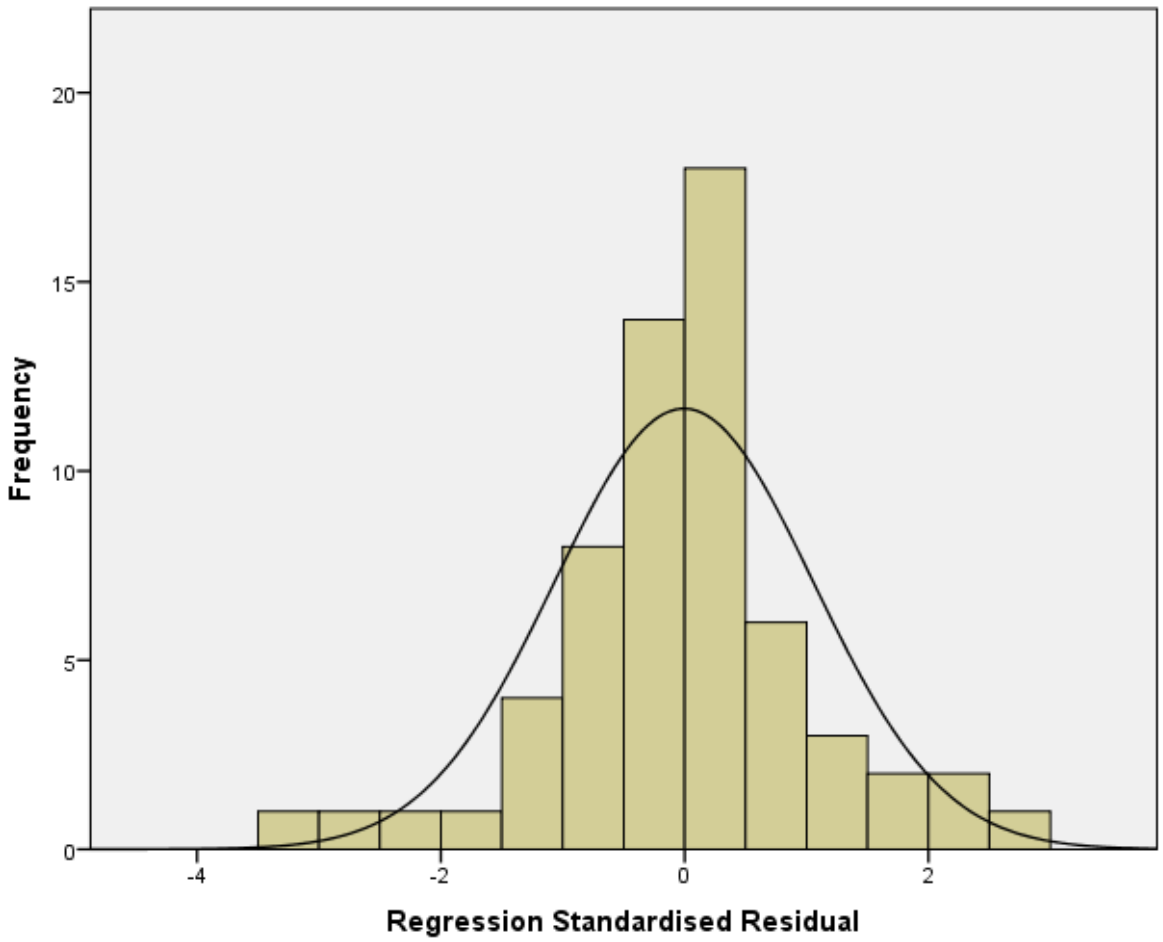
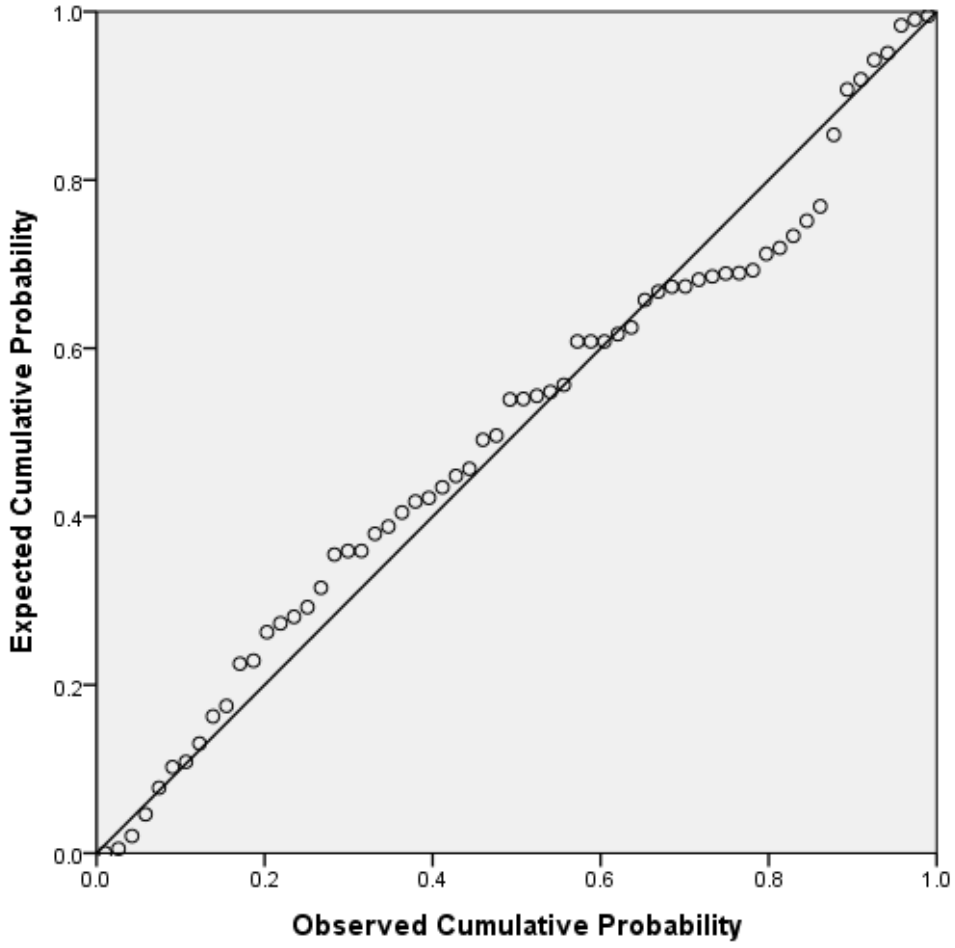


Figure 6-8: PP plot of the residuals



The close approximation of the histogram plot to a normal distribution, along with the close approximation of the PP plot to the reference line confirm the approximately normal distribution of the residuals.

6.11 Summary of findings

In considering the results pertaining to the prognostic cohort study, this chapter presented firstly the number of patients contacted for the study, those subsequently recruited to the study and those excluded or lost to follow up. As part of this, no statistically significant difference was identified between those who did and did not consent to participate in the study.

The baseline characteristics of the 76 subjects eligible and consented to the study were then presented. Notable descriptive findings included the high mean BMI, the higher proportion of females than males and the predominance of those educated to college or university level and undertaking professional, managerial or office based work. In addition, referral by GP was the predominant referral route, although there was wide variation in the treatment and imaging received prior to referral to physiotherapy. For the majority of subjects they reported this episode of shoulder pain as being their first and as being either unknown, insidious or gradual onset.

Descriptive analysis of the other candidate categories was then presented, including the high degree of variability within the data, alongside the methodological rationale for excluding the 'painful arc' variable. Descriptive analysis of the treatment delivered, period under treatment and number of appointments was detailed, highlighting a high degree of heterogeneity in the intervention.

Descriptive analysis of the outcome (OSS) was then detailed, including the categorised change using OSS MCIC from baseline to discharge, baseline to 3 months post-discharge and the more complex pattern from baseline to discharge to 3 months post-discharge. The numerical change in OSS data demonstrated a mean improvement in pain and function from baseline to discharge and baseline to 3 months post-discharge.

The logical, theoretical and statistical procedures for trimming the number of candidate variables to align with 10 cases per variable was then presented. This included the requirement to perform additional data reduction due to the smaller than intended final sample size. For the seven candidate variables, descriptive analysis of any difference in potential prognostic variables between those lost to follow up and those available for analysis was performed. This identified that a disproportionate number of those unavailable for the 3 months post-discharge model were younger patients and had lower total SPADI scores (lower pain and disability) at baseline.

The regression analysis for each prognostic model identified that when considering the change in OSS from baseline to discharge, total SPADI and Age were the only two variables that predicted

outcome, collectively accounting for 15.7% of the variance in OSS change. Directionally, a higher total SPADI (greater pain and disability) was associated with a greater improvement in OSS, whilst greater age was associated with less improvement in OSS. For the change in OSS from baseline to 3 months post-discharge, only total SPADI was predictive of outcome, accounting for 9.6% of the variance in OSS change. Directionally, a higher total SPADI (greater pain and disability) was associated with a greater improvement in OSS.

7 DISCUSSION

7.1 Overview

The aim of this study was to identify baseline factors that predicted outcome in SIS/RCTendinopathy patients following treatment with physiotherapy. Following a review of the literature in this area, five major categories of potential prognostic variables were identified which spanned three conceptual frameworks relating to SIS/RCTendinopathy. Seventy-six patients met the inclusion and exclusion criteria and received pragmatic physiotherapy treatment as determined by their treating clinician.

Data was available from 73 participants at discharge and 62 at 3 months post discharge. Using a sequential decision making process of clinical, conceptual and statistical methods, the number of potential candidate variables was aligned with the sample size. Two, multivariate regression models were constructed where the dependent variable was change in the OSS from baseline to discharge; and baseline to 3 months post-discharge.

The major findings of this study was that total SPADI and age were identified as predictors of outcome at discharge and together accounted for 15.7% of the variability in the OSS change score at discharge. Regarding the change from baseline to 3 months post discharge, only the total SPADI was identified as being a predictor and accounted for 9.6% of the variability in the dependent variable.

In discussing these findings, this chapter is structured such that firstly the methodological decisions taken and the non-prognostic findings will be explored. The order of this will mirror that of the literature review and will cover the study design and sample through to the prognostic variables considered, intervention, outcome and statistical analyses. This will lead onto a discussion of the hypotheses accepted and rejected by the current study along with the research and clinical implications. Limitations and future areas of research will conclude the chapter.

7.2 Methodological elements

7.2.1 Nature of the study design

Linking with the assessment criteria applied for prognostic studies (NHMRC 2000), the current study was a prospective cohort study. The use of two different types of hospital setting (Regional centre and DGH type setting) and a pragmatic, clinician-led approach to treatment enhance the representativeness of the findings in respect to routine practice in the NHS.

The use of follow up time periods defined according to the point of actual discharge from physiotherapy bring further clinical-relevance to the study findings. In spite of the pragmatic and highly variable treatment period, retention rates were above 95% at discharge and above the threshold of 80% at 3 months post-discharge. These compare favourably with the published literature and mean that the current study sits in the highest category of prospective cohort prognostic study grading (Phillips et al. 2016).

Embedding the study within routine clinical practice had multiple research benefits such as environmental validity and minimal inconvenience for subjects. However it also brought logistical challenges including the necessity to avoid contamination of the data by exposure to therapeutic elements and data collection being influenced by pressures on clinical service provision. This second aspect likely had a negative impact on the size of the final sample.

The posting of subjective questionnaires to patients prior to their first appointment had the advantage of capturing their state prior to any therapeutic intervention, including physiotherapy assessment. But despite written requests to only complete these questionnaires close to the time of their appointment, uniformity in the time between completing the questionnaires and attending could not be guaranteed. Conversely, aspects such as VAS were completed face to face and so consistency of timing was assured.

As per the methods chapter, sequencing of the physical testing was performed such that activities that might be influenced by pain but were of limited likelihood to exacerbate symptoms were undertaken first (i.e. ROM assessment) whilst those that were almost certain to exacerbate symptoms such as resisted strength testing and orthopaedic symptom reproduction tests were undertaken later. The ultrasound scan was performed last because aside from being able to attain optimal imaging positions, the imaging findings were highly unlikely to be influenced by acute symptom aggravation. As such, the complete battery of subjective, physical and imaging data was collected in a single visit, which immediately preceded the patient's first physiotherapy assessment. It was felt that such an approach minimised likely order effects and balanced scientific rigour with the demands of an NHS based study undertaken by a single researcher.

7.2.2 Nature of the sample

7.2.2.1 Sample size and nature of those who did and did not consent

The target sample size (n=100) was determined by the number of potential prognostic variables and the guide of 10 cases per variable (Peduzzi et al. 1996). However due to study time-pressures and the post-recruitment exclusion of eight subjects from the study, the final sample size of 76 was smaller than intended. Trimming of the number of candidate variables was undertaken to reduce the likelihood of over-fitting of the subsequent prognostic model. However this inevitably led to a narrower range of variables being considered due to the reduction in statistical power, thereby limiting the scope of the study. A larger research team would be one mechanism to help address this.

In the context of previously published papers in this area it is noted that the current study had a higher number of cases per variable than Hung et al. (2010) with 0.5, Engebretsen et al. (2010) with 5.5, Mintken et al. (2010) with 6 and Tyler et al. (2010) with 7; and comparable to Bartolozzi et al. (1994) and Kennedy et al. (2006) with 10 and Chester et al. (2016) with 11.8. As such, the current study is able to provide results that likely can be replicated in a new sample. However it is acknowledged that some authors (Concato et al. 1993) advocate the number of cases with a *positive outcome* as the mechanism for defining the number of candidate variables. Were this adhered to then using the OSS MCIC, only 5 variables would have been included in the current study as those subjects that 'improved' were n=51 and n=47 at discharge and 3 months post-discharge (respectively). This would have further reduced the number of candidate variables. However given the anchoring of this study at the exploratory end of the prognostic framework proposed by Kent et al. (2010), the use of the total number of cases per variable was deemed permissible.

No statistically significant difference in age or gender was identified between those who consented to the study and those who did not. Furthermore although the large number of those not consented for whom referral source was not recorded (n=25) does raise questions as to the robustness of this data, the source of recorded referral was very similar in both groups. Such findings indicate that systematic bias in the recruitment of subjects was unlikely to be present for the variables considered.

7.2.2.2 Pathology (including inclusion and exclusion)

The current study applied a novel approach to identifying an SIS/RCTendinopathy sample whereby clinical presentation suggestive of the target pathology, combined with exclusion of discrete non-

SIS/RCTendinopathy pathologies, plus the clinical opinion of the subsequent treating clinician was used. This reflected the limited utility of orthopaedic tests and the poorly defined nature of the pathology in terms of both aetiology and signs and symptoms (Calis et al. 2000; Lewis 2009). As such there are some fundamental differences between the sample used in the current study and those investigations that either considered a general shoulder pain sample or alternatively identified their sample using orthopaedic test findings.

Indeed the use in the current study of a battery of orthopaedic tests with a cut-off point of 3 or more positive tests (Michener et al. 2009) as an inclusion criterion would have resulted in exclusion of 20 subjects (26% of the cohort) who were otherwise deemed eligible and consequently recruited to the study. This raises the spectre that a percentage of patients with SIS/RCTendinopathy (as identified by clinical presentation and confirmed by subsequent treating clinician assessment) are potentially being excluded from research where a battery of orthopaedic tests as an inclusion criterion is used.

In the current study, four patients were subsequently excluded as their treating clinicians identified them as having non-SIS/RCTendinopathy pathologies. The diagnoses of adhesive capsulitis as the dominant pathology (n=2) and cervical referral of the symptoms (n=2) highlights the limitations of the clinical application of inclusion and exclusion criteria at baseline. This poses a fundamental question in relation to research in this area, namely how can a sufficiently large sample be efficiently identified in a reproducible yet clinically meaningful manner. It is also noted that almost 20% of the cohort complained of pins and needles which is surprising given the exclusion of cervical referral of symptoms; potential explanation for this includes the presence of local neuropathy.

7.2.2.3 Patient care pathway

In relation to the patient care pathway, all subjects in the current study were recruited at the same clinical referral point, namely referral for physiotherapy in a secondary care setting. Whilst a total of seven patients were referred via Orthopaedics, Rheumatology and A&E, the overwhelming majority were referred by their GP, reflecting a primary into secondary care bias. The largest single route into physiotherapy via GP referral was where no treatment was administered (n=32). However in just over half of cases (n=41; 53.9%) their GP had applied one or more treatment approach (advice, oral or topical analgesia or NSAIDs, injection) before subsequently referring to physiotherapy. This introduces inherent variability into the sample because n=32 subjects had received no treatment, whilst n=41 had arguably failed conservative treatment via their GP. It also reflects substantial differences in care pathway composition, which reflects uncertainty as to the optimal primary care

management approach for SIS/RCTendinopathy, despite the publication of guidelines (Diercks et al. 2014) in this area.

Of further note is that whilst half of the cohort had no imaging prior to their physiotherapy appointment, the remaining 38 subjects had one or more of U/S, X-ray and MRI. Such data reflects substantial variation in the use of imaging within the same sample of patients, likely reflecting uncertainty as to the merit and timing or sequencing of imaging in this condition. As with the approach to primary care management, the imaging data highlights heterogeneity in the SIS/RCTendinopathy care pathway. This likely represents substantial inefficiencies in the use of clinical resources and a potentially negative contribution to chronicity and patient outcomes. The improved defining of care pathways is therefore proposed as a priority area. It would allow for subsequent prognostic research to be nested within standardised assessment and treatment routes, thereby providing more meaningful information amid a reduction in the inherent background noise of current care.

7.2.2.4 Non-prognostic variables: patient demographics and clinical history data

A large number of variables relating to demographics and clinical history were collected. However, in order to provide a focused discussion of the non-prognostic patient demographics and clinical history information, only data that might potentially influence the methods or findings in this study will be considered here.

7.2.2.4.1 BMI

The mean BMI (29.0 kgm^{-2}) is at the upper end of the 'Overweight' banding (25.0 to 29.9 kgm^{-2}) (Viester et al. 2013) and includes a large number of patients in the sample who were clinically obese. Alongside evidence of an association between BMI and shoulder pain (Vikari-Juntura et al (2008), increased adipose tissue will also inevitably impact upon both scapula-humeral assessment and ultrasound imaging. Specifically the reduction in ease of bony landmark observation and attenuation of the ultrasound signal (respectively) will impair the collection of accurate data. As both measurement tools are already noted for their poor reproducibility then elevated BMI levels are a likely additional source of error, thereby limiting the prognostic utility of these variables.

7.2.2.4.2 Gender

A higher proportion of females (59.2%) than males comprised those who were eligible and consented to participate in the study. Reasons for this are unclear but evidence from Kennedy et al.

(2006), Deutscher et al. (2009) and Chester et al. (2016) of female gender consistently predicting poorer outcome provides limited evidence that the recruited sample may have been biased towards a less favourable outcome and by the reported higher pain sensitivity in this gender (Kindler et al. 2011; Horn et al. 2014).

7.2.2.4.3 Education level and type of paid work

Over 80% of those recruited to the study had a college or university education. Although education level data were not available for those not recruited to the study, census data (http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/census/2011-census-analysis/local-area-analysis-of-qualifications-across-england-and-wales/rpt---local-area-analysis-of-qualifications-across-england-and-wales.html ; last accessed 17/02/2017) for Wales indicated 26.5% of the population as having a degree level or above qualification. Whilst direct comparison of these data are limited by methodological and geographical differences, they provided limited evidence that those recruited to the study were disproportionately well educated compared to the general population.

Potential reasons behind this may include those with a higher level of educational qualification having greater awareness of the altruistic benefit of contributing to evidence based-healthcare or greater background knowledge leading them to recognise the potential advantage of receiving a diagnostic ultrasound scan in parallel to their physiotherapy treatment. Were this second aspect the case then the inherent 'treatment effect' of receiving a diagnostic scan should be noted as a potential confounding factor, thereby limiting transferability of the findings to routine NHS care where point of care diagnostic ultrasound scanning is not typically available. Research evidence of distrust of medical research among certain groups (Cobb et al. 2014) may also help to explain such findings.

In parallel to this, Engebretsen et al. (2010) identified that lower education status (as defined by not being educated to college or university level) was predictive of worse pain and disability via absolute SPADI score and 'not working' status at 12 months post baseline. This might indicate a likely outcome advantage for the sample in the current study due to the low incidence of lower education status.

Partial triangulation with the education level data comes from the 'type of paid work' data whereby there was a much higher number in professional, managerial or office based work (n=42) compared

to manual, semi-manual or unskilled workers (n=8). Although speculative, this might reflect a perception by potential subjects that participation in the study was ominous due to the volume of paperwork (invitation letter, self-screening and then envelope of questionnaires) sent when inviting patients to participate. In spite of the use of lay language, this could indicate that improved accessibility of studies to less educated patient groups is warranted. Another aspect of the disproportionate mix of workers relates to the potential that unskilled workers may be more likely to have zero hour contracts / an unsecure job, aligning with the participation component / environmental factor of the ICF. Whilst evidence of this was not collected as part of the study, should this be the case then the higher number in professional, managerial or office based work may reflect improved opportunities to attend the NHS for the majority of the sample.

In terms of habitual use of the shoulder, demographic data demonstrated low levels of loaded and/or above shoulder height working in the cohort. This might be postulated to lead to low levels of loading of tissues within the shoulder, linking with the patho-anatomical model. However this would potentially be cancelled out by the evidence that the majority of participants played overhead sports. Analysis of demographic data to consider the nature of the sample is therefore advocated in future studies so as to identify any bias within the sample and areas for emphasis regarding how best to ensure future samples are truly representative.

7.2.2.4.4 Previous shoulder pain

In the majority of cases (80.3%) it was the patient's first episode of shoulder pain that they were presenting with. In light of the highly recurrent nature of shoulder pain (Luime et al. 2004), it is conceivable that this is not representative of the wider shoulder pain and SIS/RCTendinopathy population. Potential reasons behind this include those who have recurrent symptoms may not present to NHS medical services (as they feel the NHS cannot help them further) or are not referred onto physiotherapy (for similar reasons) or have already escalated along the care pathway to potentially more invasive services such as Orthopaedics. However it is also possible that this represents a recruitment bias where those presenting with recurrent symptoms were disproportionately less likely to choose to participate in the study. The finding from Engebretsen et al. (2010) that previous shoulder pain was predictive of worse pain and disability via absolute SPADI score at 12 months post baseline provides limited evidence that the recruited sample may have been skewed towards a more favourable outcome.

7.3 Exploration of the potential prognostic variables in terms of non-regression related findings

The selection of variables of interest, robustness of data collection procedures, data handling and variable trimming along with the descriptive and reliability study findings will now be considered for each variable category. Due to the limited methodological elements involved in the collection of the demographics and clinical history related variables, these will be considered together.

7.3.1 A. Demographics and B. Clinical history

In considering demographic and clinical history variables there are a large number of potential but disparate variables across a variety of characteristics which could be explored as prognostic factors. Where sample sizes are sufficiently large (e.g. Deutscher et al. 2009; Chester et al. 2016), multiple such variables can be considered. However where sample size necessitates limiting the number of candidate variables then informed trimming of the number of variables is necessary.

Age and duration of symptoms were chosen as the A. demographics and B. clinical history variables to be entered into the prognostic model, as detailed previously. Aligning with prognostic research best practice, age was retained as a continuous variable so as to retain maximal statistical power (Royston et al. 2009). However duration of symptoms was recorded as an ordinal variable in an attempt to address potential recall inaccuracy. This highlights one of the limitations of such variables in that they are typically patient reported and so potentially incorrect interpretation or inaccurate reporting by patients can be confounding factors. Supporting evidence for this comes from the identification by almost 2/3 of the sample in the current study that their symptom onset was insidious. As such, accurately identifying duration of symptoms will be inherently challenging where onset is gradual.

In this area of research, duration of symptoms was the most consistently identified predictor in the published literature thereby providing a degree of confidence in selecting this variable over other clinical history variables. With regards to demographic variables, the choice was less clear-cut and this highlights the issue of a large amount of potentially relevant demographic and clinical history data not being included in the prognostic model. As such, the building of a consensus as to which demographic and clinical history variables influence outcome has merit. Similarly, the use of larger sample sizes so as to incorporate more candidate variables can be advocated. However mechanisms to achieve this in an economically and logistically sustainable manner, whilst preserving clinical relevance, need to be developed.

7.3.2 C1. Patient reported measures: Pain

In the current study the request for patients to rate their momentary pain 'At rest' using the prompt "when you are just resting with your arm by your side" was used to maximise the likelihood that scores were provided in a consistent manner across the sample and with minimal impact of recall bias. The rating of momentary pain 'On activity' using the prompt "*your* painful movement" sought to optimise this patient and condition-specific score. As such this approach can be considered to address some of the shortcomings identified in the pain score approaches used by Kennedy et al. (2006), Engebretsen et al. (2010), Kromer et al. (2014) and Chester et al. (2016).

The recording of pain 'At night over the last week' was a novel element when considering the previous studies critiqued in this area of research. The relevance of night time pain has been highlighted in those with rotator cuff disease (Minns Lowe et al. 2014) thereby reinforcing the relevance of capturing this aspect of symptomatic shoulder disorders. The combining of the three pain scores into a mean value mirrored the approach taken by Mintken et al. (2010) and provided a mechanism for combining each of these clinically meaningful values. Mintken et al. (2010) arbitrarily used only the mean value. However in the current study the identification of weak to strong correlations between each of the four pain variables provided a rationale for selecting a single, conceptually meaningful value for potential inclusion in the prognostic model as a mechanism for minimising over-fitting by reducing the number of candidate variables.

However due to the limited sample size, exploration of options to exclude 3 of the 10 candidate variables was undertaken and led to the VAS pain scores being excluded due to their high correlation with the total SPADI. This reflects the conceptual overlap between the VAS pain scores and the five 'pain scale' questions on the total SPADI.

For future research in this area the finding from this study that the total SPADI was highly correlated with the mean of the three clinically relevant VAS readings – each of which were highly correlated with the mean value – provides a (partial) rationale for omitting VAS readings and using only the total SPADI. In order to address issues of over-fitting and to reduce the administrative burden upon researchers and patients, this finding warrants future implementation.

7.3.3 C2. Patient reported measures: Psychological symptoms

The variety of tools used in previous prognostic studies to explore psychological symptoms reflects the complexity of this variable. In order to capture the full breadth of psychological symptoms the 4DSQ was chosen and published cut-off points used to guide the clinical interpretation of levels of different psychological symptoms. These revealed low levels of 'strongly elevated' symptoms in the sample, from 1.4% for somatisation to 6.8% for distress. These data align with both Kromer et al. (2014) and Chester et al. (2016) who also identified low levels of psychological pathology in their shoulder pain patient samples. In light of evidence (Pincus et al. 2002) of a high incidence of psychological symptoms in people with MSK disorders, there is the potential that the current sample and those in other studies were non-representative. Potential reasons for this include those with psychological symptoms being less likely to seek help with musculoskeletal disorders or being less likely to consider participating in clinical research.

The use of the 4DSQ to identify 'moderately elevated' or 'strongly elevated' symptoms for each subsection provides a guide for treatment. This is a notable strength of the 4DSQ whereby 'moderately elevated' levels indicates possible problems, which should be monitored whilst the 'strongly elevated' category requires further diagnosis and treatment (Koorevaar et al. 2016). In dichotomising the 4DSQ data, the subsequent variable entered into the prognostic model represented the presence of one or more psychological symptoms at a level requiring monitoring or treatment. As such the psychological symptom data can be considered to have been collected in a more robust manner than that of previous studies in this area. Nonetheless the potential for patients to have incorrectly interpreted the questions or insufficiently self-identified their own symptoms are caveats to consider when considering the reported level of psychological symptoms.

7.3.4 C3. Patient reported measures: Function / Disability

Measures of self-reported function and disability have the potential to capture a wide range of shoulder pain related symptoms and impact (Payne and Michener 2014). The use in the current study of a robust, region specific measure (i.e. the SPADI) provided a mechanism for capturing this.

As with many patient reported scores, the SPADI comprises sections, each with a different emphasis (namely pain and disability) which can be analysed separately. In order to explore options for variable trimming, the current study identified high correlations between each of the components and the overall score, thereby providing a statistically justified rationale for using the total SPADI score. The subsequent identification that this score correlated closely with both the mean of the

pain VAS scores and the physical ROM variable likely reflects their close conceptual relationship with the component pain and disability SPADI scores (respectively). Such findings provide a rationale for considering patient reported measures of function and disability such as the SPADI a high priority variable to collect in prognostic studies.

7.3.5 D1. Clinical measures: Strength

The use in the current study of a HHD for quantifying strength provided a mechanism for bringing accuracy to the measuring of strength in the clinical setting whilst still ensuring direct transferability of the findings. Conversion of the force readings to a rotatory moment provided a mechanism for accommodating individual variation in lever arm length and the pre-methods study results provided evidence of excellent intra-rater reliability. These results were highly comparable to those reported in the comparable study by Awatani et al. (2016) who reported ICC's between 0.850 and 0.980. Such elements align with prognostic study recommendations (Riley et al. 2013) for inclusion of robustly collected variables and methods of recording.

The recording of IR and ER strength aligned with their relevance to rotator cuff strength deficits and strengthening (Reinold et al. 2004). However measuring these in neutral shoulder abduction positions arguably limits their clinical and functional relevance because symptoms and strength deficits typically occur towards mid-range positions (Khan et al. 2013). The use of a composite planar movement (scaption) reflected the measurement of strength in a functional plane (Harrington et al. 2013). However the close overlap of this testing position and the empty/full can test position highlights the potential threat to the subsequent strength readings from pain inhibition. Indeed this highlights some of the likely contributing factors to strength deficits.

The use in the current study of an absolute strength reading could be considered inferior to a reading normalised to the opposite side. However the typical finding of less strength on the non-dominant side – and therefore arm dominance to be a confounding factor (Westrick et al. 2013) where normalisation to the asymptomatic side is used – is illustrated in the finding that n=40 (52.6%) of subjects reported that their symptomatic side was their non-dominant arm. As such, this supports the decision to use the absolute strength readings on the symptomatic side.

In selecting the variable to be entered into the prognostic modelling, high correlation between each of the strength measures provided a justification for choosing a single measure as each can be

representative. This provides a rationale for using just a single strength test in future studies, thereby reducing burden upon the researcher and patient.

7.3.6 D2. Clinical measures: ROM

ROM was quantified using a digital inclinometer and a guided but planar movement so as to balance reproducibility and clinical relevance (Kolber et al. 2012). The recording of ROM relative to both the limit of ROM and the point of symptom exacerbation by movement ensured that any impact of these factors were captured. Furthermore, the pre-methods study results provided evidence of excellent intra-rater reliability, thereby aligning with prognostic study recommendations (Riley et al. 2013) for inclusion of robustly collected variables and methods of recording. In addition the intra-rater reliability results were highly comparable to those reported in the comparable study by Mullaney et al. (2010) who reported ICC's between 0.940 and 0.98.

However the subsequent identification that 10 subjects (13%) had no meaningful reading for the variable of 'point of symptom exacerbation' because of floor and ceiling effects raised questions about the value of this variable. Indeed the high correlation between the readings for limit of ROM and the point of symptom exacerbation by movement provides a rationale for recording only the limit of ROM data in subsequent studies, thereby reducing the burden upon the researcher and patient. However the ROM data also highlights the substantial variability in symptom behaviour in this NHS population.

The ROM variable was ultimately excluded from the prognostic modelling stage due to the limited sample size. Whilst this may be considered a substantial failing in the current study, the correlation between this variable and the total SPADI score provided a statistically justified (partial) rationale for omitting the ROM variable. Indeed the disability component of the total SPADI (comprising over 60% of the total score) has substantial conceptual overlap with the amount of movement available at the shoulder. This provides evidence for the use of total SPADI as a surrogate for the physical measure of ROM in future studies, thereby reducing researcher burden and potentially increasing sample size within a given timeframe or research budget. As such, the association between the ROM variable and the total SPADI also provides evidence to support a theory that patients are accurate in self-identifying their functional and ROM impairments.

7.3.7 D3. Clinical measures: Scapular movement and control

The choice to use the SDT in the current study was informed by the non-mutual exclusivity of movement patterns in this tool and the ability to use it in the clinical setting, thereby maximising transferability. However the SDT was developed on high-level athletes using data from loaded trials where shoulder flexion and abduction were both assessed.

Data from the current study that 42% (n=32) of the cohort were unable to perform a loaded trial provides evidence that in its original form the SDT cannot be uniformly used with NHS populations such as in the current study, due to the high levels of symptom provocation from elevating the arm whilst holding a load. Furthermore the statistically significant correlation between the gradings from the non-loaded and loaded trials provides evidence that the performance of a loaded trial in an NHS population could be deemed as not necessary.

The inter-rater reliability study provided evidence of inferior agreement in the current study (Kappa = 0.395 and 0.329; un-weighted and linear-weighted Kappa respectively) compared to the original study by McClure et al. (2009) (videotape rating: Right side Kappa = 0.61, left side Kappa = 0.48; both weighted Kappa). It should be noted that subjects in the study by McClure et al. (2009) were athletes from National Collegiate Athletic Association Divisions I and III who participated in sports requiring intense overhead arm use. Although no descriptive ROM or BMI data were reported, exclusion criteria in their study included a body mass index ≥ 30.0. In addition, the athletes were drawn from water polo, swimming, baseball/softball, volleyball and tennis. They were therefore highly likely to have much greater ROM than the subjects in the current inter-rater reliability study (mean=138°, SD=25°, min=74°). Furthermore the BMI in the current inter-rater reliability study (mean=29 kgm^{-2}, SD=6 kgm^{-2}) revealed a much higher BMI than in the study by McClure et al. (2009).

This is of relevance because some of the components of normal scapulohumeral motion and scapular dyskinesis detailed in the tool that comprises the SDT are typically seen in the mid to higher ranges of shoulder elevation (Ettinger et al. 2014). As such those patients in the current study with substantially impaired active ROM may have been unable to attain the range of movement where some elements of either normal scapulohumeral motion or scapular dyskinesis occur. Categorising of such patients may have been an additional source of disagreement between raters, thereby contributing to the lower Kappa.

The higher BMI in the current inter-rater reliability study might conceivably make it harder to accurately observe the bony landmarks making disagreement between raters more likely. As such these data provide evidence that the SDT may not be sufficiently robust to use in NHS populations. In light of prognostic research study recommendations regarding robustness of variables (Riley et al. 2013), it raises the question of whether scapular movement and control are too difficult to accurately quantify for them to be considered in a prognostic study. The use of a composite scapular variable (Lewis 2009) which emphasises symptom reduction may be an alternative approach.

7.3.8 E1. Structural pathology via imaging

The recording of a broad spectrum of pathology related features using high-resolution sonography is a novel aspect of this study when considering the previous studies critiqued in this area of research. Furthermore the undertaking of this by a researcher whose original qualification is that of a treating clinician (physiotherapist) mirrors the recent shift towards point of care-based imaging in musculoskeletal disorders (Roll et al. 2016). The emphasis here was on gaining a full clinical qualification rather than simply project specific training, thereby reflecting the long learning curve for shoulder ultrasound (McDonald et al. 2010; Beggs 2011). The subsequent publication (Smith et al. 2015) stemmed from the challenges that MS faced during his training and represents an attempt to facilitate the high quality training of those undertaking shoulder ultrasound imaging.

The inter-rater reliability findings demonstrated a range of agreement levels from fair (Kappa of 0.298 and 0.358 for calcific deposits and cuff pathology; respectively) to moderate (Kappa of 0.503 and 0.477 for bursitis and PTT; respectively) to almost perfect agreement (0.841 for FTT) (Viera and Garrett 2005). However the large number of missing datasets for bursitis and the lack of statistically significant agreement for tendinopathy findings means that the internal validity and reliability of the variable entered into the prognostic model (no pathology; 1 of bursal or cuff pathology; both bursal and cuff pathology) must be viewed with caution.

The inter-professional agreement levels in the current study are broadly comparable to those reported by Thoomes-de Graaf et al. (2014) who also compared sonography findings of Physiotherapists to Radiologists. Specifically they reported a Kappa of 0.54 for bursitis and 0.28 for calcific deposits (0.50 and 0.30 in the current study; respectively). However there was wider variation between the studies for PTT and FTT with a Kappa of 0.10 and 0.63 (0.48 and 0.84 in the current study; respectively).

Possible reasons for this include the training of MS under PM and therefore the increased likelihood of comparable scanning technique and image interpretation. The use by Thoomes-de Graaf et al. (2014) of a general shoulder pain sample mirrored that of the current inter-rater reliability study and the percentage incidence of Radiologist reported pathology in the study by Thoomes-de Graaf et al. (2014) and the current inter-rater reliability studies were almost identical. Furthermore the primary care based Thoomes-de Graaf et al. (2014) study mirrored the largely GP referred sample (n=30; 86%) in the current inter-rater reliability study, making these unlikely reasons for the observed differences in reliability values. Nonetheless, the reproducibility of any diagnostic ultrasound variables must be optimised for it to be included in future prognostic studies. Mechanisms to achieve this include using experienced Sonographers and employing mechanisms to improve diagnostic reproducibility such as standardised terminology.

As noted in the methods section, for the main cohort study MS employed specific mechanisms (bilateral scans to provide a pseudo comparator and use of specific criteria relating to the sonographic appearance of cuff tendinopathy and bursitis) in an attempt to improve the robustness of these sonographic findings which had non-significant agreement and low incidence of PM reported findings; respectively. These mechanisms have yet to be formally assessed for reliability or acceptability but recent publications in this area (Ingwersen et al. 2016; McCreesh et al. 2016) support the use of standardised scanning procedures and definitions in order to improve inter and intra-rater reliability for tendinopathy-related findings.

The collapsing of pathology findings in various portions of the rotator cuff into a single binary rating (e.g. calcific deposits: yes/no) does not reflect diagnostic ultrasound scanning procedures where pathology is identified specifically in relation to the different portions of the rotator cuff (Jacobson 2011). The subsequent collapsing of all pathology findings into a three level variable (i.e. no pathology; one of bursal or cuff pathology; both bursal and cuff pathology) takes this one step further. When considering the potential differences in the management of calcific deposits or tendinopathy *versus* PTT or FTT (Baring et al. 2007) then this approach arguably has limited clinical transferability. However this compromise in variable trimming approach had the advantage of retaining low incidence pathologies (PTT and FTT had an incidence in the main cohort of 2.6% and 3.9%; respectively) in the final analysis and also reducing the number of variables to be analysed.

7.3.9 E2. Structural pathology via orthopaedic tests

The potential inclusion of orthopaedic test results as a prognostic factor but not as an inclusion criterion is a novel aspect of this study. As such it reflects the increasingly held opinion that orthopaedic tests are of limited diagnostic utility (Lewis 2009; Kelly et al. 2010; Alqunaee et al. 2012; Hanchard et al. 2013) but nonetheless recognise a potential role for predicting outcome. As such, this approach aligns with the emphasis within the prognostic methodology literature of overcoming the limitations of diagnosis via prognostic research (Hemingway et al. 2013).

The omission of the painful arc variable from the results was in part due to threats to the conceptual value of the variable, an element of which was the large portion (n=26; 34%) of the sample whose limited active ROM would have precluded them having a meaningful painful arc finding. This highlights one of the challenges of assessing a general NHS population where a variety of symptom severity and functional limitation can be present. The finding also raises questions regarding the merit and clinical utility of the painful arc in such a population.

The collapsing of the remaining four test results into a single variable (yes/no: 3 or more positive test results) reflects published evidence regarding the limited diagnostic utility of *individual* tests and simultaneously assisted with reducing the number of candidate variables. As noted earlier in the discussion chapter, just over a quarter of subjects had 2 or fewer positive test results and so conceivably would have been excluded from the study using a criteria of 3 or more positive tests as a cut-off point (Michener et al. 2009) as in the prognostic study by Hung et al. (2010). However the approach taken in the current study involved consideration of the clinical presentation and was supplemented by the treating clinicians who identified any subjects who as treatment continued were identified as being non-SIS/RCTendinopathy. This double-screening procedure is arguably a more clinically meaningful, albeit time intensive approach to subject recruitment. However this also reflects the challenges of identifying a homogenous SIS/RCTendinopathy patient sample, which reflects this difficult to define pathology (Linaker and Walker-Bone 2015). Indeed the necessity for larger sample sizes when researching in this area poses significant logistical challenges when applying this time intensive, clinically meaningful screening process. The alternative approach of recruiting generic shoulder pain samples provides one mechanism to overcome this, as evidenced by the larger scale studies by Deutscher et al. (2009), Sindhu et al. (2012) and Chester et al. (2016).

The use of three or more positive tests as a cut-off point (Michener et al. 2009) facilitated the reducing of the number of candidate variables but in an evidence informed manner. However due to

the limited sample size the orthopaedic test variable was not entered into the prognostic model phase. Structural pathology as a variable was still retained for modelling via the diagnostic ultrasound data and this reflected published evidence of its superior sensitivity and specificity (Iagnocco et al. 2003; Micheroli et al. 2015). Nonetheless, orthopaedic tests are easier to apply in routine clinical practice and so the omitting of this variable is a serious consequence of the limited sample size. As such it removes the opportunity to consider structural pathology as a potential prognostic factor in routine outpatient physiotherapy where immediate access to ultrasound imaging would likely be prohibitive.

7.4 Treatment / intervention – contrasted with other shoulder prognostic studies

7.4.1 Measurement tool

The bespoke treatment recording tool used in this study was developed in collaboration with the treating clinicians at the two data collection sites. In the subsequent prognostic study, clinicians had the opportunity to record any interventions they delivered which were not included on the form, but in all cases these did not yield additional data. All categories were used at least once indicating that there was no redundancy in the tool. Combined with the refinement of the initial version based on clinician feedback this provides limited evidence of the content validity of the tool (Mori et al. 2016).

The covering sheet (appendix 4; first page of 'Final version of proforma for collecting patient treatment information') reminded the treating clinician that treatment data analysis would be anonymised and non-judgmental. However it is possible that clinicians may have consciously or unconsciously altered their treatment approach or recording of treatments due to being observed. The impact of the so-called 'Hawthorne effect' (Kohli et al. 2009) must therefore be acknowledged as a potential confounding factor.

The treatment approach in the current study contrasts with those papers that used standardised interventions including those that were derived from RCTs where treatment was defined a priori. However where the intervention involved a degree of examination-based treatment adaptation such as with Virta et al. (2009) and Kromer et al. (2014) then comparison with the clinician-determined intervention in the current study can be more readily made.

There is potentially greater comparability with the entirely clinician directed approach used by Deutscher et al. (2009) although review of their categories demonstrates only a limited overlap with those used in the current study. This perhaps reflects differences in clinical approach or professional

autonomy between the UK and Israel (where the Deutscher et al. (2009) study was undertaken). Furthermore the treatments used by Deutscher et al. (2009) were across all shoulder diagnostic categories. Comparability might also be assumed between the work of Chester et al. (2016) and the current study as both were UK NHS based and treatment was entirely clinician directed. However Chester et al. (2016) did not publish details of the treatments delivered nor of how the delivered intervention was recorded. As such, it is not possible to compare the treatment received by patients in their study and the current study.

The approach taken in the current study provides a potential foundation for the routine capture of what constitutes clinician directed care of patients with SIS/RCTendinopathy. As such, it could form a template for prospective testing of whether sub-groups of patients respond to different interventions, so called treatment effect modifiers (Kent et al. 2010). However this requires prospective testing in a larger cohort alongside formal testing of the inter- and intra-rater reliability, inter-departmental reliability and criterion validity of the treatment recording tool.

7.4.2 Treatment delivered

The treatment was delivered by band 6 and 7 clinicians for the majority of patients (n=69; 91%) reflecting the guidance given to the administrative staff to allocate patients in the study to these clinician bands. This was to control for the potential confounding factor of the early learning curve of newly qualified staff as an additional source of treatment variability (Macznik et al. 2015). Whilst a greater consistency of treatment intervention might be expected from using experienced clinicians, transferability of the findings to routine practice where there is a wide range of skill mix would be limited. In the current study, six patients were subsequently treated by band 5 clinicians and this reflects the necessary pragmatic approach with a clinical based prognostic study.

The high frequency with which education and advice, exercise therapy and home exercise programmes were used as the treatments delivered mirrors the consistent use of these modalities in the literature (Dong et al. 2015) and provides partial evidence that the treatment delivered was representative of typical physiotherapy. Despite published evidence of some benefit from combining manual mobilisation with exercise (Green et al. 2003; Desjardins-Charbonneau et al. 2015; Steuri et al. 2017), hands-on techniques such as manual therapy were only applied in the treatment of half of the cohort whilst taping and electrotherapy were rarely utilised. This may reflect the individual patient presentations or clinical treatment preferences. However the NHS based nature of the study and the likely pressure for patient throughput may influence the predominance of hands-off

techniques. Transferability of the findings to a private funded clinical setting where such circumstances may be reversed may therefore be limited.

As half of the cohort in the current study were not treated with manual therapy, then comparability with the prognostic study by Mintken et al. (2010) will inevitably be limited as those authors used only spinal manipulations as their intervention. Conversely the frequent use of exercise therapy in the form of stretches, strengthening, stability and ROM exercises in the current study broadly mirrors the approaches used by Conroy and Hayes (1998), Deutscher et al. (2009), Virta et al. (2009), Engebretsen et al. (2010), Hung et al. (2010) and Kromer et al. (2014), so comparison of outcome with these studies can be more readily undertaken.

The high frequency of a home exercise programme mirrors that of Chester et al. (2016) who reported its use for 99% of participants. The optimising of treatment carry over and encouragement of patient independence and empowerment likely reflects the NHS based treatment investigated in the current study and by Chester et al. (2016). However it should be noted that no attempt was made in the current study to quantify adherence to HEP and so the reported levels of prescribing HEP may be an overestimate of the actual HEP performed by the cohort.

It must also be acknowledged that unlike routine clinical practice, all subjects in the current study received a diagnostic ultrasound scan. Patients may conceivably have derived a therapeutic impact by implied confirmation or exclusion of structural pathology, which they might have deemed relevant to their symptoms. Clinicians may have been similarly influenced in their treatment approaches. These potentially confounding or even therapeutic factors merit formalised investigation but were beyond the scope of the current study.

7.4.3 Treatment period and discharge status

A large number of prognostic studies in this area used standardised duration of treatment and numbers of appointments, reflecting their RCT or standardised intervention approaches. Typically these were much more intensive than the current study, for example with Kromer et al. (2010) treatment was twice weekly for 5 weeks. Conversely, the mean of 4.6 appointments over 13 weeks in the current study likely reflects the service pressures of routine clinical care and subsequently a less intensive but more protracted treatment phase. Such differences have the potential to impact on the subsequent outcomes and the timescales for assessing change.

The number of appointments in the current study (mean = 4.6, SD = 2.6) was less than those in the pragmatic studies of Deutscher et al. (2009) (mean = 9, SD = 6) and the Canada based study by Kennedy et al. (2006) (mean = 15, SD = 9) but largely comparable to that of Chester et al. (2016) (mean = 5.0, SD = 2.7). This may reflect international differences in the location of the clinical settings along with local funding pressures (private *versus* public health). As such the NHS based collection by Chester et al. (2016) and the current study likely reflect the availability of healthcare resources in the NHS and illustrate the caution that must be applied when comparing findings from different healthcare systems. It should also be noted that the quoted figures for Chester et al. (2016) were those who provided follow up data and so may be skewed by loss to follow up, whilst in the current study the number of appointments data was for all subjects.

The treatment period is also of relevance, but not only from the perspective of treatment 'dose'. In this respect Kennedy et al. (2006) and Deutscher et al. (2009) reported treatment lasting a mean of 65 days (SD=26) and 56 days (SD=40); respectively. These were shorter and less variable then the mean of 13 weeks (91 days) and SD of 10 weeks (70 days) in the current study. The high degree of variability in the treatment duration in the current study indicates a highly heterogeneous intervention. However it also highlights that where studies such as Kromer et al. (2014) undertook their follow up at 3 months, many subjects (were they receiving a pragmatic intervention) would theoretically still be under treatment; specifically 47% (n=36) in the current study. For the study most comparable to the current NHS based study – Chester et al. (2016) where their follow up occurred at 6 weeks and 6 months after the initial appointment – many subjects would theoretically still be under treatment. Specifically 77% (n=58) of subjects in the current study were still under treatment after 6 weeks, and 16% (n=12) were still under treatment after 6 months. Assuming greater comparability across these two NHS based studies, the conceptual relevance of follow up based on arbitrary time points is questionable. The use of 'discharge from physiotherapy' as the reference time point in the current study brings the benefit of clinical relevance but in so doing introduces further variability (i.e. background statistical noise) into the study and might provide a partial explanation for the limited power of the final models to predict outcome. From a logistical perspective, protracted treatment phases also led to delays in the final data collection and analysis.

The patient status at discharge in the current study provided evidence that the majority (n=57; 75%) of those recruited to the study completed physiotherapy treatment and were subsequently discharged. This is broadly comparable to the segregated findings of the most comparable study

(Chester et al. 2016) who reported 61% and 87% of their cohort completing physiotherapy (those who didn't provide follow up outcome data and those that did; respectively).

In the current study, the 13 patients (17%) who were discharged subsequent to DNA or UTA represent wasted clinical resources. Given the pressures on valuable NHS resources the predicting of discharge secondary to DNA or UTA could be considered a high priority area of research and one where prognostic modelling might be able to play a valuable role.

The discharge for referral onto orthopaedics for surgery consultation (n=2), injection via orthopaedics (n=1) or GP (n=1) and specialist psychological treatment (n=1) is evidence of variability in the SIS/RCTendinopathy care pathway. The rate (7%) of referral-on to other specialities is likely to be artificially low as such data was only collected at discharge and so will not capture post-discharge utilisation of other portions of the SIS/RCTendinopathy care pathway. An extended longitudinal follow up of this data type is therefore advocated. Along with two patients who received a shoulder injection from their GP whilst still under physiotherapy treatment, the referral-on numbers reveal a complex web of patient care pathways that likely comprise routine practice. This is both an additional source of inherent variability in the study of SIS/RCTendinopathy and potentially a contributing factor to patient care inefficiencies.

7.5 Exploration of outcome variable

7.5.1 Outcome variable and data handling (OSS)

As noted previously, several prognostic studies (Bartolozzi et al. (1994), Kennedy et al. (2006), Deutscher et al. (2009), Engebretsen et al. (2010), Kromer et al. (2014) and Chester et al. (2016)) used the same patient reported score as their dependent variable and as a potential prognostic factor. In order to partially address the methodological issues posed by this, the current study used a different self-reported function and disability score (SPADI) as the determinant of outcome (via the OSS).

The published close correlation of the OSS with the SPADI and DASH (Dawson et al. 2001; Cloke et al. 2005; van der Linde et al. 2015) indicates that the OSS measures similar constructs to those studies that used the SPADI (Engebretsen et al. 2010; Kromer et al. 2014; Chester et al. 2016), DASH (Kennedy et al. 2006) and QuickDASH (Chester et al. 2016). Consequently a degree of direct comparability between the current study and those that used similar outcome measures is permissible. However the published close correlation of the OSS with the SPADI indicates that the

use in the current study of the SPADI as a potential prognostic factor and the OSS as the determinant of outcome is unlikely to substantially address the above concerns regarding overlap between patient reported scores as the dependent variable and as a potential prognostic factor.

This highlights a challenge for future prognostic research in this area, namely how best to retain patient reported outcome of disability / function as a potential prognostic variable (given the overwhelming evidence in the literature of its consistent prognostic relevance) *and* use a clinically relevant measure of outcome (essentially patient reported outcome of disability / function) but whilst avoiding the methodological overlap.

7.5.2 Data handling using the OSS MCIC

The MCIC from Ekeberg et al. (2010b) was used to identify change in status and this provided a mechanism for establishing the clinical response of the cohort. However it must be acknowledged that the Ekeberg et al. (2010b) MCIC data was derived from a 1-week recall period rather than the prescribed 4 week period described by the OSS developers (Dawson et al. 1996) and used in the current study. As such, the change in status data in the current study must be viewed with caution.

The OSS MCIC scores demonstrated for those subjects from whom follow up data were available that from baseline to discharge 51 subjects (69.9%) improved, four subjects (5.4%) worsened and 18 subjects (24.7%) remained the same. For the period baseline to 3 months post-discharge 47 subjects (75.8%) improved, three subjects (3.9%) worsened and 12 subjects (15.8%) remained the same.

The multi-stage baseline to discharge to 3-month post-discharge data demonstrated that the largest group were those who improved from baseline to discharge and maintained this relative to baseline at 3 months post-discharge. However there was a complex pattern of change across the three time points as illustrated by Figure 6.2: “Representation of change in OSS over time for each subject”. This included those who improved and then further improved (n=4; 6.5%), remained the same from baseline at each time point (n=5; 8.1%), were the same at discharge but at 3 month post-discharge had improved from baseline (n=8; 12.9%) and those that improved from baseline to discharge but were within the MCIC of baseline (i.e. same relative to baseline) at 3 month post-discharge (n=5; 8.1%). Such data highlights the complexities of multiple time points and the individualised pattern of change demonstrated across the cohort.

As no control group was used it is not possible to specify cause and effect or to differentiate out natural course of the condition. Nonetheless the rates of 'improvement' (69.9% for baseline to discharge and 75.8% for baseline to 3 months post-discharge) in the current study are comparable to the non-RCT studies which reported rates of successful outcome, namely Mintken et al. (2010) (61%), Morrison et al. (1997) (67%), Hung et al. (2010) (70%) and (Ogon et al. 2009) (73%). This provides limited evidence that a high proportion of patients with SIS/RCTendinopathy demonstrate improvement following physiotherapy and that this is maintained at 3 months post discharge.

Furthermore only a small proportion of patients worsened over this period (5.4% for baseline to discharge and 3.9% for baseline to 3 months post-discharge) in the current study which are comparable to those previously reported of 3% (Engebretsen et al. 2010), 4% (Mintken et al. 2010) and 7% (Kennedy et al. 2006). No data was specifically collected in the current study regarding adverse effects, although from the treatment collection forms and discussion with clinicians no reporting of adverse effects occurred. As such, the study provides additional, if limited, data of the benefit – with low levels of negative impact – of physiotherapy in patients with SIS/RCTendinopathy.

7.5.3 Differences between those lost to follow up and those where follow up data was available

Any systematic differences in the baseline characteristics (considered in the prognostic modelling stage) of those who were lost to follow up might have the potential to influence the prognostic models calculated on those who were available for follow up. Although only analysed descriptively, no clear evidence of a systematic bias was evident in relation to baseline OSS, symptom duration, psychological symptoms, strength, scapular movement and control or ultrasound evidence of pathology.

However there was the potential for a systematic bias whereby a disproportionate number of those unavailable for the 3 months post-discharge model were younger patients. Evidence from Deutscher et al. (2009), Kennedy et al. (2006) and Morrison et al. (1997) of older age being associated with poorer absolute outcome status can be combined with evidence from Kennedy et al. (2006) of younger age being predictive of greater improvement in disability via change in DASH score from baseline to discharge / 12 weeks. As such there is some evidence that the sample available for the discharge to 3 month post-discharge modelling phase were less likely to improve due to a loss of younger patients. However any conclusions as to an impact on the constructed models is limited by

the multi-variate statistical interactions in prognostic modelling and also the absence of age as a predictor in the discharge to 3 month post-discharge model.

In relation to total SPADI score there was limited evidence of a systematic bias whereby a disproportionate number of those unavailable for the 3 months post-discharge model had lower total SPADI scores and so had lower pain and disability scores at baseline. Evidence of greater functional impairment / disability at baseline predicting a worse outcome state (Bartolozzi et al. 1994; Kennedy et al. 2006; Engebretsen et al. 2010; Chester et al. 2016) supports a viewpoint that those available for the 3 months post-discharge follow up point may have been less likely to improve. Evidence from Kennedy et al. (2006), Hung et al. (2010) and Kromer et al. (2014) of higher baseline disability predicting greater improvement in status suggests that those lost to follow up at 3 months may have influenced the subsequent 3 month model which total SPADI was retained in.

It is interesting to note that the 2 variables (age and total SPADI) where a systematic bias was identified in those unavailable for the 3 months post-discharge model were the same variables that comprised the baseline to discharge prognostic model (age and total SPADI) and the baseline to 3 months post-discharge prognostic model (total SPADI). A statistical explanation for this is unclear, but it highlights the importance of analysing loss to follow up datasets.

7.5.4 Timing

The collection of the OSS relative to discharge from physiotherapy arguably optimises the clinical meaningfulness of the resulting prognostic findings. However due to the pragmatic nature of the study this time point was highly variable relative to when physiotherapy commenced. This had the effect of increasing the timescales for completion of the study and also introduced an additional source of variability into the study design. However it provides some evidence that studies such as Kromer et al. (2014) and Chester et al. (2016) which collected outcome data at arbitrary time points may have been pooling data from those who had already completed treatment with those who would be under treatment for a considerably longer period. Thus the prognostic data reported by Kromer et al. (2014) and Chester et al. (2016) arguably comprises an uncertain mix of patients at very different points in their clinical care.

The use of meaningfully spaced follow up time points has merit in providing evidence regarding stability of change and prognostic value beyond the point of discharge. However the use by Mintken et al. (2010) of follow up occurring a matter of days after baseline data collection makes their data of

very limited clinical value. One merit though of the approach of Mintken et al. (2010) is that potential confounding factors such as the natural course of the condition, variability in adherence to home exercise programmes and exposure to potentially aggravating work, recreational or functional tasks is minimised. As such, the current study and those who included longer follow up time periods have this as both potential confounding factors and potential sources of inherent background noise within their study designs. However as these factors all represent inevitable real world elements they are inherent within longitudinal study designs and can either be quantified by attempting to measure their occurrence, or sufficiently large sample sizes used to ensure a sufficiently powered study to accommodate them. One mechanism by which some of the inherent variability (regarding timescales) could be managed would be to use a framework that applies a maximum duration of treatment.

7.6 Statistical analysis: Analysis approach and data reduction procedures

7.6.1 Data handling: prognostic modelling

The use of a pre-defined, stratified approach to trimming the number of candidate variables provided a mechanism for aligning the number of variables with the sample size. This approach to avoiding over-fitting and potential under powering of the study aligns with prognostic methodology recommendations (Riley et al. 2013) regarding statistical analysis. However as noted earlier, basing the number of variables on the number of cases with a positive outcome is recommended by some authors.

The use of clinical and logical methods for trimming the number of variables could be argued to potentially introduce bias into the analysis approach. However such an approach ensures maximal clinical relevance of the subsequent findings. Furthermore, an over-reliance on purely statistical criteria can have the disadvantage of producing spurious results. The subsequent use of a combination of clinical, logical and statistical methods in the current study is therefore a constructive attempt to address the shortcomings of both polarised approaches. It is acknowledged that the subsequent removal of the VAS, ROM and Orthopaedic test data from the final analysis could be considered un-ethical given the patient and researcher time and cost required to collect these variables. However it is argued that the importance of avoiding over-fitting – and the potential for inaccurate clinical prediction rules to be generated from such data – would be a more un-ethical and harmful outcome.

The retention of the OSS as a continuous variable had the advantage of maximising statistical power in the study. The subsequent regression model was constructed using an automated, statistically determined procedure of forward selection and backward elimination with pre-assigned statistical cut off points. The use of a stepwise but non-hierarchical approach allowed all permutations of variables to be considered irrespective of their category, which ensured that no single category or variable was prioritised. However, it must be acknowledged that the retention a minimum of one variable per category and of both the diagnostic ultrasound and scapular movement and control variables (irrespective of issues with their reproducibility) could be construed as use of a pseudo-hierarchical approach.

The subsequent application of thorough testing of fit of the model and diagnostics ensured that the model for each time point could be viewed with a degree of confidence. As part of this the thorough reporting of R^2 values (indicating the amount of variation in the outcome variable explained by each model) and regression equation coefficients (β) (used to identify the direction of any relationships between predictor variables and the dependent variable) enables the reader to ascertain the magnitude and nature of the relationships identified by the analysis.

7.7 Summary of methodological considerations

It can therefore be seen that the current study with its prospective cohort design and greater than 80% follow up at each time point makes it of the highest level of prognostic study (NHMRC 2000; Phillips et al. 2016). The combination of logical, theoretical and statistical procedures for aligning the number of candidate variables to the sample size provides a multi-factorial mechanism for addressing potential threats of over-fitting. The subsequent use of a quantified, pragmatic treatment and follow up points defined by the actual point when face-to-face treatment finished all ensure maximal clinical relevance of the study.

The inclusion of the SDT and diagnostic ultrasound are note-worthy novel elements. Although their reproducibility were both lower than preferred for inclusion in a prognostic model, the analysis of their robustness provides a mechanism for considering this when interpreting the overall study findings. Alongside these the inherent variability in the nature and dose of the intervention (physiotherapy) combined with heterogeneity in the care pathway within which the study was located are all identifiable sources of background noise which may contribute to the low level of predictive ability.

7.8 Hypotheses accepted / rejected

The following null hypotheses were **accepted** and the corresponding experimental hypotheses rejected:

B.(i) Clinical history

In a multivariate analysis, clinical history related variable will not be a predictor of outcome between baseline and discharge in SIS/RCTendinopathy patients treated with physiotherapy.

D.(i) Clinical measures

In a multivariate analysis, baseline clinical measures will not be a predictor of outcome between baseline and discharge in SIS/RCTendinopathy patients treated with physiotherapy.

E.(i) Structural pathology

In a multivariate analysis, baseline structural pathology will not be a predictor of outcome between baseline and discharge in SIS/RCTendinopathy patients treated with physiotherapy.

A.(ii) Demographics

In a multivariate analysis, demographics will not be a predictor of outcome between baseline and 3 months post-discharge in SIS/RCTendinopathy patients treated with physiotherapy.

B.(ii) Clinical history

In a multivariate analysis, clinical history related will not be a predictor of outcome between baseline and 3 months post-discharge in SIS/RCTendinopathy patients treated with physiotherapy.

D.(ii) Clinical measures

In a multivariate analysis, baseline clinical measures will not be a predictor of outcome between baseline and 3 months post-discharge in SIS/RCTendinopathy patients treated with physiotherapy.

E.(ii) Structural pathology

In a multivariate analysis, baseline structural pathology will not be a predictor of outcome between baseline and 3 months post-discharge in SIS/RCTendinopathy patients treated with physiotherapy.

F.(ii) Multivariate prognosis

In a multivariate analysis, a combination of 2 or more of A. Demographics, B. Clinical history, C. Patient reported measures, D. Clinical measures and E. Structural pathology will not predict outcome between baseline and 3 months post-discharge in SIS/RCTendinopathy patients treated with physiotherapy.

The following null hypotheses were **rejected** and the corresponding experimental hypotheses accepted:

A.(i) Demographics
In a multivariate analysis, demographics will not be a predictor of outcome between baseline and discharge in SIS/RCTendinopathy patients treated with physiotherapy; specifically Age (when combined with mean SPADI) was identified as accounting for 6.7% of the variability in the dependent variable. The negative coefficient meant that greater age was associated with less improvement in OSS.

C.(i) Patient reported measures
In a multivariate analysis, baseline patient reported measures will not be a predictor of outcome between baseline and discharge in SIS/RCTendinopathy patients treated with physiotherapy; specifically mean SPADI was identified as accounting for 9% of the variability in the dependent variable. The positive coefficient meant that a higher SPADI mean (greater pain and disability) was associated with a greater improvement in OSS.

F.(i) Multivariate prognosis
In a multivariate analysis, a combination of 2 or more of A. Demographics, B. Clinical history, C. Patient reported measures, D. Clinical measures and E. Structural pathology will not predict outcome between baseline and discharge in SIS/RCTendinopathy patients treated with physiotherapy; specifically mean SPADI and Age were identified as accounting for 15.7% of the variability in the dependent variable.

C.(ii) Patient reported measures
In a multivariate analysis, baseline patient reported measures will not be a predictor of outcome between baseline and 3 months post-discharge in SIS/RCTendinopathy patients treated with physiotherapy; specifically mean SPADI was identified as accounting for 9.6% of the variability in the

dependent variable. The positive coefficient meant that a higher SPADI mean (greater pain and disability) was associated with a greater improvement in OSS.

7.9 Exploration of prognostic findings in relation to previously published prognostic studies

In this section the prognostic findings from the current study will be considered in the context of other studies in this area and in the realm of prognostic research more broadly.

7.9.1 Hypotheses A.(i) and A.(ii) Demographics

In rejecting the null hypothesis and accepting the experimental hypothesis, the current study found in multivariate analysis that the demographics factor of age was a predictor of change in OSS between baseline and discharge but not between baseline and 3 months post-discharge. Directionally, the current study identified that greater age was associated with less improvement in OSS.

Morrison et al. (1997), Kennedy et al. (2006) and Deutscher et al. (2009) all reported greater age to be predictive of poorer outcome although their findings referred to absolute score at discharge. Of greater comparability to the current study was the finding from Kennedy et al. (2006) that younger age was predictive of greater improvement in disability via change in DASH score from baseline to discharge / 12 weeks. From these results it could be concluded that greater age predicts both greater impairment at follow up and less improvement in impairment at follow up. Such findings mirror those from the low back pain literature with respect to improved pain (Verkerk et al. 2015) and greater gains in perceived function (Gregg et al. 2014) for younger patients and poorer change in disability in older patients (Cecchi et al. 2014).

Whilst the current study and Morrison et al. (1997) dealt specifically with SIS/RCTendinopathy patients, both Deutscher et al. (2009) and Kennedy et al. (2006) considered less well defined shoulder patient populations, namely shoulder pain and soft tissue disorders of the shoulder; respectively. Along with the findings from the low back pain rehabilitation literature, this provides evidence that the conceptually generic variable of age may be of relevance in predicting outcome across different, albeit overlapping patient populations.

It is acknowledged that eight other shoulder prognostic studies failed to identify age as a predictor and that many of these were of a high quality. However in the case of Hung et al. (2010), the age of

their subjects (20-33 years) provides a very narrow range of age data for any association to be identified and thus could explain the lack of age being a predictor. Where studies dichotomised the age variable (Bartolozzi et al. 1994; Mintken et al. 2010) or used very small sample sizes (Conroy and Hayes 1998) the greatly reduced statistical power might also explain the lack of significant findings.

Timescale might also be another explanatory factor because the current study only identified age as a predictor of change at the discharge time point but not at 3 months post discharge. This would align with those who had null findings for age at a longer time period, namely 6 months (Ogon et al. 2009), 12 months (Engebretsen et al. 2010) and 20 months (Bartolozzi et al. 1994). Such findings indicate that whilst increasing age may influence outcome in the short term, any influence is lost at longer follow up time points. Speculative reasons for this include an initial delay in response to intervention in those who are older, perhaps due to delayed tissue healing (where a patho-anatomical model is favoured) or a reduced ability to accept and integrate adaptive activities into their lives where a psychosocial model is favoured (Bjorck-van Dijken et al. 2008) or both, where an ICF model is favoured.

In terms of an emerging consensus on predicting outcome, the current study adds to the evidence base that age is a predictor whereby older age is associated with worse outcome. However the large number of studies which did not find age to be predictive, together with the wide variety of demographic variables which were found by a small number of studies to be predictive, means that a definitive consensus cannot currently be established. However it is recommended that future prognostic studies in this area retain age as a continuous variable for consideration in their prognostic model.

7.9.2 Hypotheses B.(i) and B.(ii) Clinical history measures

In accepting the null hypotheses, the current study did not find in multivariate analysis that clinical history related measures were a predictor of change in OSS between baseline and discharge nor between baseline and 3 months post-discharge. Specifically the duration of symptoms prior to baseline data collection was not a predictor of outcome.

This finding is in agreement with that of Conroy and Hayes (1998), Ogon et al. (2009), Virta et al. (2009) and Engebretsen et al. (2010). However these previously published studies were generally of inferior quality compared to the three studies that found duration of symptoms to be predictive of outcome state and four studies that found it predictive of greater change.

A potential explanation for these differing results (non-predictive *versus* predictive) lies in how the duration variable was treated. In this respect some of those who found this variable to be predictive (Deutscher et al. 2009; Mintken et al. 2010; Bartolozzi et al. 1994) dichotomised chronicity to acute *versus* chronic, greater *versus* less than 90 days and greater *versus* less than 12 months; respectively. This is in contrast to the use of ordinal duration of symptom data in the current study. However Kennedy et al. (2006) also used ordinal data for this variable and did find it to be predictive of change in disability from baseline to discharge. Therefore differences in the processing of this variable are unlikely to explain the conflicting findings.

Given that a majority of studies in this research area and also many from other musculoskeletal disorders have found duration of symptoms to be predictive (Rihn et al. 2011; McClinton et al. 2015) then specific elements pertaining to the current study may provide an explanation. In this respect the frequency of patients reporting this as being their first episode of shoulder pain (80.3%) is unusually high given the high recurrence rate of shoulder pain (Winters et al. 1999). This may therefore reflect a skewed sample in terms of those being referred to physiotherapy and/or being recruited to the study.

The clinical history data also provide detail on the context of the current study, which may again present challenges when comparing findings with those studies considered in the literature review, and also those studies excluded. The literature reviewed in this thesis was limited whereby those studies where the intervention was steroid injection +/- non-invasive, multimodal physiotherapy and where results were not segregated for those patients who were treated solely with non-invasive, multimodal physiotherapy were not considered. Yet for pragmatic reasons 16% of the cohort in the current study had received one or more injections via their GP as part of their episode of care, thereby potentially tainting the sample with this additional modality. It is therefore possible that studies excluded because they did not present segregated results for non-invasive, multimodal physiotherapy separate from injection and non-invasive, multimodal physiotherapy would actually have provided highly relevant findings.

It can therefore be concluded that the current study does not add to the emerging consensus of duration of symptoms being predictive of outcome. However the clinical history data provides some differences in patient sample, which may limit comparability of the current study findings with previous work in this area. The lack of consensus regarding the other clinical history variables

explored in this area of research indicates that further work is required to establish the predictive value of other variables.

7.9.3 Hypotheses C.(i) and C.(ii) Patient reported measures

In rejecting the null hypothesis and accepting the experimental hypothesis, the current study found in multivariate analysis that patient reported measures were a predictor of change in OSS between baseline and discharge and between baseline and 3 months post-discharge. Specifically function as defined by the total SPADI predicted outcome at both time points. However the patient reported measure of pain as measured by VAS was excluded from the final prognostic analysis due to the limited sample size and the emphasis upon avoiding over-fitting of the model. Psychological symptoms as measured by the 4DSQ were not found to be predictive at either time point.

7.9.3.1 Pain

The previous literature in this area provided moderate evidence of baseline pain levels being a predictor, with higher pain levels predicting worse outcome. The current study is unable to directly contribute to this debate due to the omission of the pain variable from the prognostic modelling stage. As noted previously this was permissible due to the high correlation with the patient-reported functional outcome score (total SPADI).

7.9.3.2 Psychological

The finding from the current study that psychological symptoms as measured by the 4DSQ were not predictive contrasts with that of Engebretsen et al. (2010), Sindhu et al. (2012) and Chester et al. (2016). However unlike the 4DSQ, the patient expectation predictors (Chester et al. 2016) and general health status via EQ-VAS (Engebretsen et al. 2010) provide limited evidence regarding actual psychological symptoms.

Where the more relevant measures of fear-avoidance beliefs (Sindhu et al. 2012) and pain self-efficacy (Chester et al. 2016) were identified as predictive of outcome this was in studies with sample sizes greater than 1,000. Conversely, the same factors were not predictive in the studies by Mintken et al. (2010) and Kromer et al. (2014); Engebretsen et al. (2010), who had sample sizes of 80, 90 and 104 (respectively). Where more formal assessments of psychological symptoms were collected as in the study by Engebretsen et al. (2010) and the current study, these were not found to be predictive. Of note is that they also had sample sizes a factor of 10 smaller than those of Sindhu et al. (2012) and Chester et al. (2016).

These elements provide a potential explanation regarding the apparently conflicting psychological symptoms findings, which is that whilst the psychosocial model is advocated in musculoskeletal disorders and non-invasive, multimodal physiotherapy, its actual impact is small. A lack of impact from treatments targeting psychosocial factors in similar conditions like low back pain (Jellema et al. 2005) provides a degree of support for this viewpoint, as does evidence of a lack of association between distress and rotator cuff tendinopathy (Mallows et al. 2017). As such the findings raise the possibility that the complex nature of psychological symptoms make them inherently difficult to identify as predictors, thereby requiring large scale, sufficiently powered studies in order to identify a predictive effect.

Furthermore, the limited incidence of clinical levels of psychological distress in the current cohort provides partial triangulation for this, as does the low psychological scores in the cohort of Kromer et al. (2014) and Chester et al. (2016). One mechanism to explore this would be the inclusion of a category within the treatment measurement asking for specific psychological interventions, e.g. counselling and exposure therapy. However the addressing of psychological elements that is argued to be inherent within the patient-therapist interaction (Dragesund and Kvale 2016) make this a challenging aspect of treatment to capture.

7.9.3.3 Function

In the current study, baseline function / disability as measured by the total SPADI score was predictive of change in OSS from baseline to both discharge and 3 months post-discharge. The identification of this variable as a predictor agrees with the majority of studies (7 out of 8) that explored it as a potential predictor. As such, there is strong evidence from the literature that baseline function / disability predicts outcome and this aligns with published systematic reviews (Littlewood et al. 2013).

This consensus might be explained by the generic importance of function, i.e. an aspect that is consistently of relevance to patients. This theory would be supported by the broad range of components that comprise patient reported function scores (e.g. pain, mobility, ability to participate) and this fitting with the multi-factorial nature of musculoskeletal conditions, including SIS/RCTendinopathy (Walker-Bone et al. 2003). It could also be perceived as reflecting the complex intervention nature of non-invasive, multimodal physiotherapy (as demonstrated by the multi-

faceted treatment intervention data in the current study) and the capacity to potentially influence multiple facets of the condition (Littlewood et al. 2016).

However there is also a potential methodological explanation for these findings in that the aspects of function / disability measured at baseline will likely have substantial overlap with the aspects of function / disability measured at follow up, including the derivation of change scores. Such a scenario is highlighted by the use of the same measurement tool in both capacities in the studies by Kennedy et al. (2006) with the DASH; Chester et al. (2016) with the Quick DASH; Engebretsen et al. (2010), Kromer et al. (2014) and Chester et al (2016) with the SPADI; Bartolozzi et al. (1994) with the UCLA; Deutscher et al. (2009) with the CAT. This situation could be termed 'outcome alignment', whereby the consideration of a potential predictor which is conceptually (or in the above cases actually) the same as the dependent variable will more likely give rise to statistically significant findings. As such, the use in the current study of two different but conceptually overlapping measures (SPADI and OSS) partially mirrors this methodological explanation.

Of further note is that the total SPADI score was the only variable in the current study to be predictive at both time points. This finding mirrors that of Chester et al. (2016) who found both the SPADI and Quick DASH to be predictive at both the 6 week and 6 month time points. As such, these pooled findings can be viewed as providing further evidence of the consistent predictive ability of this variable type.

Directionally, a higher SPADI mean score (greater pain and disability) in the current study was associated with a greater improvement in OSS. This concurs with the result from Hung et al. (2010) whereby higher baseline disability as measured by the FLEX-SF was predictive of improvement which was a binary classification based upon the GROC score. These also concur with the reporting by Kromer et al. (2014) of higher baseline disability levels as measured by the function element of the SPADI as being predictive of a greater change score at 3 months post baseline. Furthermore Kennedy et al. (2006) found worse physical health (via the Physical Component Score (PCS) of the SF-36) at baseline to be predictive of greater improvement in disability via change from DASH score at baseline to discharge / 12 weeks.

Indeed, of the studies that explored function / disability as potential predictors and where change in status was the dependent variable, only Mintken et al. (2010) did not find the variable to be predictive. However the very short study duration from Mintken et al. (2010) of approximately 1

week (from baseline to follow up) makes it possible that such established elements of pathology as measured by the SPADI would be unlikely to change. In support of this is the weighting of the SPADI whereby eight of the 13 questions pertain to disability.

However the predictive direction in the current study appears to be opposite to that reported by Bartolozzi et al. (1994), Kennedy et al. (2006), Deutscher et al. (2009), Engebretsen et al. (2010) and Chester et al. (2016), whereby greater functional impairment / disability at baseline predicted a worse outcome state. The seemingly opposite findings between the studies (including the current one) whose dependent variable was change in state versus those whose dependent variable was absolute state might be explained by the way in which outcome was defined.

The explanation here could be that Bartolozzi et al. (1994), Kennedy et al. (2006), Deutscher et al. (2009), Engebretsen et al. (2010) and Chester et al. (2016) identified that those with high levels of disability at baseline were those who still had high levels of disability (relative to the cohort) at the follow up time point. Such a finding need not be incompatible with the findings of the current study, Kennedy et al. (2006), Hung et al. (2010) and Kromer et al. (2014), where those with higher baseline disability demonstrated the most change, potentially because they had the most scope to improve. From a statistical perspective, the regression modelling in the current study and similar investigations would have been strongly influenced by those with very large changes in function / disability outcome measure and would have differentiated clearly between subjects with worsening of score and improvement of score. However with Bartolozzi et al. (1994), Kennedy et al. (2006), Deutscher et al. (2009), Engebretsen et al. (2010) and Chester et al. (2016), the magnitude of change from baseline would have no impact on the regression modelling as only end state was used.

In essence it is possible for subjects to change the most in a cohort and yet still be poorer functioning / more disabled than others in the cohort. Such findings highlight the critical role played by the methods used to define the dependent variable, the challenge of pooling results from studies using different approaches to measuring outcome and consequently the challenges of establishing a consensus on prognosis in this area.

7.9.4 Hypotheses D.(i) and D.(ii) Clinical measures

In accepting the null hypothesis, the current study did not find in multivariate analysis that clinical measures were predictors of change in OSS between baseline and discharge nor between baseline and 3 months post-discharge. Specifically neither of strength (as represented by the mean moment

of external rotation on the symptomatic side) nor scapular movement and control as represented by the scapular dyskinesis test were predictors of outcome. ROM represented by the limit of ROM on the symptomatic side was excluded from the final prognostic analysis due to the limited sample size and the emphasis upon avoiding over-fitting of the model.

7.9.4.1 Strength

Hung et al. (2010) identified reduced serratus anterior strength as being predictive of improved outcome although in their study IR, ER and abduction strength were not predictive. Bartolozzi et al. (1994), Kennedy et al. (2006), Mintken et al. (2010) and Chester et al. (2016) considered various aspects of strength but none found it to be a predictor of outcome. The weight of evidence is that strength is not predictive of outcome and this is supported by the findings from the current study.

Use in the current study of an accurate and reproducible strength measurement tool addresses measurement error as a potential source of statistical background noise. Nonetheless, one possible explanation for the null finding are other sources of between-subject variability. Possible explanations include the wide variation in duration of symptoms (39.5% < 6 months duration; 15.8% >24 months duration) and thus theoretically wide variation in levels of disuse muscle atrophy (Vlaeyen and Linton 2000). Wide variation in pain levels at the time of first assessment (VAS of 0.0 to 9.4) could potentially introduce variability in pain inhibition levels (Verbunt et al. 2005). Similarly, the wide variation in baseline upper limb activity levels across the cohort (represented by demographic data of recreational and work-related activities) was another potential source of inherent strength data variation. The subsequent interaction between these factors provides an additional layer of potential complexity within this variable. Indeed these reflect the inherent complexity in the manifestation of musculoskeletal pain, including in SIS/RCTendinopathy.

One mechanism by which such sources of variability might be accommodated is by assessing subgroups of patients according to their symptom duration, symptom intensity or background levels of shoulder activity. However the generalisability of such findings would be limited.

7.9.4.2 ROM

As noted previously, the disparate nature of those ROM variables that were found to be predictive in previous studies (Mintken et al. 2010; Tyler et al. 2010; Chester et al. 2016) – along with the much larger number of ROM variables which were not found to be predictive – makes establishing a consensus on the predictive value of ROM very challenging.

In considering the current study, comparison would have been limited by the following factors:

- In reporting posterior shoulder tightness as predictive, Tyler et al. (2010) were actually investigating the change in ROM across the treatment period as the variable of interest, whilst the current study was concerned with baseline ROM as a predictor. As such, fundamentally different uses of ROM-related data were employed in the current study and that of Tyler et al. (2010).
- Mintken et al. (2010) on the other hand identified baseline ROM variables as being predictive of a better outcome. It should be noted that the sample size available for prognostic analysis in the current study at discharge (n=73) was comparable to that of Mintken et al. (2010) (n=80). However whilst the current study used a stepwise procedure to enter seven variables into the regression model, Mintken et al. (2010) entered double this number. The less conservative approach taken by Mintken et al. (2010) raises the potential for over-fitting of their model.
- The baseline active ROM variable identified by Chester et al. (2016) as being a predictor was only predictive for their end state SPADI score at 6 weeks. In the current study a pragmatic, clinician-led treatment approach was used, as was the case with Chester et al. (2016). However at 6 weeks, 77% of subjects in the current study were still under treatment. Therefore the timing of the 6-week finding by Chester et al. (2016) is likely to reflect a heterogeneous mix of patients who have completed and who are still under treatment.

Nonetheless, ROM was not entered into the prognostic model in the current study for the statistical reasons previously presented. In light of the wide-ranging aspects of ROM that have been investigated in previous prognostic studies – and predominantly found not to be predictive – there is a lack of clarity as to the predictive merit of ROM and what manifestation of it might be meaningful in predicting outcome.

7.9.4.3 Scapular movement and control

Mintken et al. (2010) considered scapular position, movement and control related variables but found them to not be a predictor of outcome and this concurs with the findings of the current study. Although there are noteworthy differences such as the use of static measures applied in the Mintken et al. (2010) study compared to the current study, the parallel use by Mintken et al. (2010) of a dynamic, qualitative approach (Kibler et al. 2002) is broadly comparable with the method used in the current study. Such findings provide a degree of triangulation with uncertainty as to the clinical

relevance of scapular movement / dyskinesis and whether people with and without shoulder pain have differences in scapular function (Timmons et al. 2012).

However Hung et al. (2010) identified less IR of the scapula as a predictor of better outcome. Specifically this was at 30° shoulder elevation during the descending phase in the unloaded condition. This reflects a key measurement difference, which was the use of a motion analysis system allowing for highly detailed measurement of 3 dimensional scapular kinematics. It should be noted however that the 60°, 90°, and 120° humeral elevation positions, the ascending phase and the other kinematic variables (upward rotation and anterior tilt) did not yield predictors. This might provide an explanation for the seemingly disparate results between the current study and that of Hung et al. (2010) in that the scapular dyskinesis test incorporates both ascending and descending phases, across all shoulder elevation ranges and provides a composite rating from all component kinematic elements. It might therefore be postulated that any such predictive scapular elements in the current study might have been dwarfed by the majority of non-predictive scapular elements.

However this potential explanation does highlight one of the key limitations of including scapular variables in a predictive study design, in that scapular movement is inherently difficult to quantify in a meaningful, comprehensive and reliable manner. It might be argued that with existing clinical tools the complexity is such that their inclusion as candidate variables in a predictive study is unlikely to yield useful results. Conversely, the use of more treatment-orientated approaches such as the symptom modification test (Lewis 2009) might yield more promising prognostic results. In addition, Hung et al. (2010) entered 48 scapular variables into their prognostic model. With a sample size of only 32 and the consideration of a further 12 other potential prognostic variables, the threat of over-fitting is particularly likely in this study, with only 0.5 cases per variable.

The use by Chester et al. (2016) of just such a symptom modification approach was novel and was identified by the authors as a predictive variable at the 6 months time point for both their end state SPADI and QuickDASH scores. Fundamental differences between this symptom adjustment approach and the movement and control approach used in the current study limits direct comparison. However the arguably more direct clinical relevance of the approach used by Chester et al. (2016) provides some limited support for the use of this scapula-related characteristic in future studies. Of further note here is that Chester et al. (2016) examined a general shoulder pain sample and so their study provides evidence of the relevance of symptom modification with scapular facilitation techniques to be independent of the type of shoulder pain pathology.

The absence of clinical measures being identified as predictors in the current study is noted. Due to the time consuming nature of collecting clinical measures then this provides one mechanism by which the requisite increasing of prognostic study sample sizes might be achieved whilst accommodating logistical and financial constraints. The conceptual overlap and statistical correlations between the total SPADI score and both ER mean moment and limit of ROM provide a rationale for selecting patient reported measures rather than clinical measures. As such they also highlight the ability for patients to fairly accurately self-identify their own levels of physical impairment.

7.9.5 Hypothesis E.(i) and E.(ii) Structural pathology

In accepting the null hypothesis, the current study did not find in multivariate analysis that structural pathology was a predictor of change in OSS between baseline and discharge nor between baseline and 3 months post-discharge. Specifically U/S findings as characterised by (i) no sonographic evidence of pathology, (ii) sonographic evidence of either bursal or rotator cuff pathology and (iii) sonographic evidence of bursal and rotator cuff pathology, were not predictors of outcome. Orthopaedic test findings were excluded from the final prognostic analysis due to the limited sample size and the emphasis upon avoiding over-fitting of the model.

The finding in the current study that structural pathology via imaging were not predictive contrasts with all of those previous studies (Bartolozzi et al. 1994; Morrison et al. 1997; Ogon et al. 2009) that have considered this variable type and subsequently found it to be predictive. However comparability with the results from Morrison et al. (1997) are very limited as the retrospective cohort study by Morrison et al. (1997) only considered acromion morphology which is of limited direct relevance to the sonographic variables considered in the current study.

The study by Bartolozzi et al. (1994) identified rotator cuff pathology in the form of moderate or large rotator cuff tears as being predictive of outcome. However it is noted that all patients in the Bartolozzi et al. (1994) study received a diagnosis of pathology ((i) impingement, tendinitis, (ii) PTT or small FTT, (iii) tears >1cm^2). This contrasts with the current study whereby n= 14 (18%) of patients were identified as having no sonographic evidence of pathology and this rose to n= 28 (37%) with a criteria of no sonographic evidence of rotator cuff pathology. Fundamental differences in the nature of the samples and/or the potential that Bartolozzi et al. (1994) inappropriately assigned an imaging rotator cuff pathology diagnosis to *all* their participants means that comparability of the different

prognostic findings in relation to structural pathology cannot be made between the current study and that of the retrospective cohort study by Bartolozzi et al. (1994).

The higher quality prospective cohort study with >80% follow up by Ogon et al. (2009) identified multiple sonographic factors as predictive of positive outcome (i.e. radiographic classification of the calcific deposits and the sonographic penetration of the deposits) and negative outcome (i.e. bilateral calcific deposits, location and volume of the deposits). Multiple fundamental differences in study design may account for these differences. Specifically the current study was concerned with SIS/RCTendinopathy whereby only n=6 (7.9%) of the cohort had sonographic evidence of calcific deposits, compared to the study by Ogon et al. (2009) where imaging confirmed presence of calcific deposits was their primary inclusion criterion. Therefore there is very limited comparability of the current cohort with that of Ogon et al. (2009) in relation to the structural pathology characteristics they explored.

Further key differences include their entering of four radiographic and three sonographic calcific variables into their regression model whilst the current study entered only one sonographic variable into the regression model whereby calcific deposits comprised only one element (cuff pathology) of one grading of the variable. This reflects the different emphases of the two studies with Ogon et al. (2009) concerned primarily with radiographic and sonographic variables whilst the current study explored variables across structural pathology, clinical measures, patient reported measures, demographics and clinical history domains. Indeed this highlights one of the quandaries of prognostic research: with Ogon et al. (2009) narrowly aligning the nature of the sample with the candidate prognostic factors to explore, this will arguably be more likely to produce prognostic models where multiple predictors are retained in the final model. Conversely the use of a more loosely defined sample with candidate prognostic factors from a broad range of aspects (as in the current study) makes it potentially more challenging to identify multiple predictors in the final model. The limited sample size upon which the prognostic models were run in the current study (n=73 and n=62 at the discharge and 3 months' time points) and subsequent limited statistical power will arguably compound this further.

Although the non-prognostic findings in relation to structural pathology align with published systematic review findings (Littlewood et al. 2013), it nonetheless can be concluded that the current study is the first prognostic study of its kind which included imaging of structural pathology but did not identify it as a predictor of outcome. The above section provides various possible explanations

for this difference, in relation to the published literature. However of note is the highly subjective nature of imaging and particularly sonographic diagnoses, which has led some authors (Moons et al. 2009) to caution against the inclusion of such variables in prognostic research. The preliminary use by the candidate of mechanisms such as bilateral scanning and criteria for the identification of tendinopathic change might be one such mechanism to improve repeatability but this remains an area for future work.

7.9.6 Orthopaedic test evidence

In the literature review it was concluded that there was weak evidence for orthopaedic tests being predictive of outcome. In the current study the exclusion of orthopaedic test findings from the final prognostic analysis means that the evidence base around this variable type is not contributed to.

To further compound this lack of clarity, the identification by Mintken et al. (2010) of a single test (negative Neer's sign) as being predictive of a positive outcome, conflicts with research opinion (Hegedus et al. 2012) that a battery of tests should be used in order to overcome issues of sensitivity and specificity with individual tests. Chester et al. (2016) did not identify the external rotation lag sign as a predictor but this test concerns rotator cuff tears (Castoldi et al. 2009).

The approach by Chester et al. (2016) of using a singular, structural integrity test is conceivably inferior to the use of diagnostic imaging in the current study. However the incumbent time and cost requirement of scanning placed inevitable limits on the sample size in the current study thereby reducing the overall power to detect prognostic factors. Furthermore, transferability to routine clinical practice is arguably better with the approach taken by Chester et al. (2016) for the same reasons.

7.9.7 Hypotheses F.(i) and F.(ii) Multivariate prognosis

In rejecting the null hypothesis and accepting the experimental hypothesis, the current study found in multivariate analysis that a combination of variables were predictors of change in OSS between baseline and discharge but not between baseline and 3 months post-discharge. Specifically total SPADI and age combined to explain 15.7% of the variance in OSS change between baseline and discharge. In the baseline to 3 months post-discharge model, total SPADI explained 9.6% of the variance in OSS.

The inclusion of only two predictor variables in the final baseline to discharge OSS model and one variable in the final baseline to 3 months post-discharge OSS model are at the lower end of the number of variables comprising multi-variate models in the previously published literature. Where studies used change of state as the outcome variable their final models comprised 2 (Kromer et al. 2014), 3 (Hung et al. 2010), 4 (Mintken et al. 2010) or 5 (Kennedy et al. 2006; Ogon et al. (2009)) variables. As with the current study these were prospective cohort studies with >80% follow up (except for Kennedy et al. (2006) and Kromer et al. (2014)) and so considered of the highest level of study quality. Study quality is therefore unlikely to explain the broad range in the number of variables comprising the final multivariate models.

Wide variation in the methods used to construct the final prognostic models provides one potential explanation for such variation. This included Kromer et al (2014) who was the only one of the above studies to retain numerous variables in their model, irrespective of the contribution or level of significance, thereby adjusting their model for these factors. The studies by Kennedy et al. (2006), Ogon et al. (2009) and Kromer et al. (2014) used an approach whereby variables with $p \geq 0.05$ were removed. Conversely Mintken et al. (2010) required variables with a $p<0.10$ so as to be retained. A combination approach was used in the current study and also by Hung et al. (2010) whereby a significance of $p \leq 0.05$ was required to enter a variable into the model and a significance of $p \geq 0.10$ was required to remove it. Evidence to support this comes from Bekkering et al. (2005) who analysed the same dataset using different regression modelling techniques and generated differing numbers of predictive variables.

The amount of variation (15.7%) explained by the baseline to discharge OSS model and for the baseline to 3 months post-discharge model (9.6%) in the current study are the lowest of any of the previously reported studies in this area. Potential explanations include the use of larger samples and therefore more highly powered studies by Chester et al. (2016) and Deutscher et al. (2009) whose final models explained 31% and 30% of the variance in their outcome variables. However studies with sample sizes that were very comparable to the current study such as Engebretsen et al. (2010) (n=104) and Kromer et al. (2014) (n=90) were able to explain 30% and 48% of the variation in their outcome variables; respectively. Such findings make it unlikely that sample size alone explains the differences in amount of variation predicted by the regression models.

Looking beyond shoulder pain, the low back pain literature provides some limited evidence of comparatively low predictive values. For example the investigation of chronic low back pain patients

by Steffens et al. (2014) was only able to predict 10 % of the variance in the 1-year pain outcome via baseline pain intensity. For their dependent variable of disability at 12 months, a combination of duration of current episode, baseline disability and educational level accounted for 15% of the variation. Interestingly the authors speculated that the partial absence in their cohort of previously identified predictors (due to the intentional skewing towards high educational and socioeconomic status patients accessing a private paying exercise program incorporating cognitive behavioural therapy) was a potential explanation for their low levels of prediction. Comparatively low levels of predictive ability in the low back pain literature were reported by Bekkering et al. (2005) where a disproportionately high number of positive responses by patients may have made discriminative prediction challenging.

However an alternative explanation for the low level of prediction in the current study may come from the multi-factorial variability within the *overall* study. This variability would conceivably create a large amount of statistical background noise, thereby making it much harder to statistically identify predictive factors which (collectively) would account for a higher percentage of variability in the outcome measure. The most notable source of variability in the current study is the use of pragmatic treatment which varied substantially between patients, including the duration, 'dose' and nature of physiotherapy. As such, a potential explanation for the even lower percentage of predictive capacity in the baseline to 3 months post-discharge model might be further variability introduced by the nature and level of shoulder activities that different patients may have undertaken once discharged from physiotherapy.

An additional source of inherent variability is the time between baseline data collection and discharge data collection. The studies critiqued in this thesis used either standardised treatment and/or standardised post-baseline data collection time points. The use in the current study therefore of individualised baseline to discharge durations means that such variability introduces further background noise into the overall study. Picking through this background noise in order to predict an adequate level of variation in the outcome variable will inevitably be more challenging and may explain the low level of variability predicted in the current study.

Potential mechanisms to address this include limiting variation within the system by partially standardising the intervention and also the follow up time. Other alternatives include aligning the nature of the population with the predictor variables as with Ogon et al. (2009) who recruited patients with calcific tendonitis and entered characteristics of the calcific deposits into their model.

From a generalisability perspective, increasing the sample size might allow for an improved capacity to predict a higher percentage of variation whilst also maximising generalisability.

7.10 Implications of findings from the current study

In the above sections an array of research and clinical implications were highlighted. In the following section, some key aspects of these will be further explored.

7.10.1 Implications in relation to SIS/RCTendinopathy conceptual frameworks

A number of conceptual frameworks within which SIS/RCTendinopathy can be viewed were presented in the literature review chapter. The prognostic findings from the current study will now be considered in relation to these frameworks.

The first of these was the patho-anatomical model, incorporating both intrinsic and extrinsic factors. The failure of the current study to identify either E. Structural pathology (intrinsic factors) or D. Clinical measures; specifically scapular movement and control (extrinsic factors) as predictors of outcome provides evidence that this model does not assist with modelling prediction of outcome in SIS/RCTendinopathy patients treated with physiotherapy.

The psychosocial model was represented by C. Patient reported measures; specifically 4DSQ which was also found to not be predictive. A role of this model in predicting outcome in this context is therefore not supported by the study findings.

Finally, the ICF classification incorporated both factors which were not found to be predictive and those that were. Specifically:

Not predictive:

Functioning and disability:

- Body function: Mental function = C. Patient reported measures; specifically 4DSQ
- Body function: Neuromusculoskeletal and movement related functions = D. Clinical measures; specifically strength and scapular movement and control
- Body structure: Structure related to movement = E. Structural pathology

Qualifier:

- Duration = B. Clinical history; specifically duration

Predictive:

Functioning and disability:

- Body function: Sensory function and Pain = C. Patient reported measures; specifically pain (represented by total SPADI score)
- Activities and participation = C. Patient reported measures; specifically SPADI

Contextual factors:

- Personal factors = A. Demographics; specifically age

Data from this study therefore provides evidence of a limited ability for conceptual frameworks to contribute to predicting outcome in SIS/RCTendinopathy patients treated with physiotherapy. Specifically only some elements of the ICF classification contributed to predicting outcome. The lack of contribution from the more narrowly focused patho-anatomical and psychosocial models can be seen as partial triangulation with the concept of SIS/RCTendinopathy being a complex, multifactorial condition. Indeed even the prediction of outcome in SIS/RCTendinopathy by the multi-faceted ICF model was only achieved by a limited portion of the framework.

This raises questions as to whether the current models of disease have utility when exploring prognosis; and whether predicting outcome in SIS/RCTendinopathy is too complex to be adequately modelled. However it has been highlighted in the prognostic methodology literature that predicting outcome is not the same as explaining cause (Moons et al. 2009). This provides an alternative explanation; namely that bespoke predictive frameworks need to be constructed. Whilst these can be informed by models of disease, they should neither be constrained by them, nor their utility judged by them.

7.10.2 Research implication

Noting the points highlighted earlier in this chapter, research implications from the current study findings include consideration of how SIS/RCTendinopathy subjects are identified. In terms of prognostic studies, the novel method of identifying the sample partly via exclusion of discrete other shoulder pathologies provided a high rate of correct pathology identification where treating clinician opinion was used as a determinant of pathology type. Specifically of the 84 patients recruited to the study, only four were subsequently deemed to be of a different primary pathology, equating to 95.2% (80/84) accuracy. This provides a novel mechanism for identifying patients in subsequent studies of SIS/RCTendinopathy cohorts. The prospective testing of the inter-rater reliability of this approach is therefore advocated.

The statistically significant correlation between the mean VAS score and the 3 constituent scores provides a rationale for using only the mean score in future studies, therefore providing a mechanism for data reduction. However the limited strength of some of these correlations must also be recognised. The same applies with the strength readings whereby the highly significant correlations between each of the strength variables supports the use of a single strength measure. However the finding that the total SPADI score was highly statistically significantly correlated with both the mean VAS score and limit of ROM variables provides a mechanism for substantially reducing both researcher and patient burden whereby the more readily collected tool of SPADI can be used instead.

The inter-rater reliability studies provided evidence that ROM and strength measures as undertaken were highly reproducible by a single rater and thereby future use in research settings can be advocated. However the same cannot be said for the SDT which in the current study was found to be conceptually flawed (regarding the use of loaded trials) and with poor levels of inter-rater agreement. An important caveat here is that these relate to NHS patients and specifically to the patients and raters involved with the current study. However they highlight the necessity of undertaking reliability studies and conceptual testing when a measurement tool is to be used in a substantially different setting to the one where it was originally developed and tested. Regardless, the conclusion from the current study is that the SDT should not be used in its current form with NHS patient populations.

Alongside the elements highlighted elsewhere in this chapter, the use of bilateral scanning as a pseudo comparator combined with a descriptive terminology to quantify the very subjective category of tendinopathy provide potential routes for the improvement of ultrasound as a measurement tool in future research. Mechanisms by which these could be investigated include testing this for reliability (within and between raters when looking at similar sonographic conclusions) and validity (outcome following treatment; also tissue biopsy for histology).

The treatment measurement tool developed for the current study provides a potential template for quantifying the component parts of physiotherapy for patients with SIS/RCTendinopathy. This could contribute to the subsequent describing and stratifying of the intervention as a mechanism for aligning treatment approaches with particular patient groups or exploring more homogenous intervention types.

The identification of patient reported outcome scores as a consistent predictor of outcome provides a rationale for making this a high priority variable to include in future research. However the risk of excessive prognostic alignment between potential prognostic factor and outcome variable must be acknowledged and as such the use of alternative or parallel determinants of outcome should be noted, e.g. patient satisfaction.

The identification of patient reported measures of function / disability along with age as predictors of outcome, provide evidence to support their inclusion in future prognostic research. When considering a method framework of prognostic research, confirmatory hypothesis-testing in an independent sample is advocated. However, due to the low percentage of variance explained by the current study it is recommended that a move from the exploratory end of study design approach cannot be advocated at this time. Further work on refining the tools to be used, or mechanisms to increase sample size, or nesting of future work within a section of the care pathway are advocated first.

7.10.2.1 'Background noise' and potential mechanisms to address this

The predictive capacity of the current study was low. In the context of this study, where potential prognostic factors across a wide range of categories were explored, where treatment was pragmatic and the timescale for follow up defined by the individualised cessation of treatment, then predictive ability was limited.

One potential explanation for this is the multi-level variability within the system. Specifically the complex, multifactorial nature of shoulder pain (including SIS/RCTendinopathy) will itself be a source of 'background noise' in terms of presentation and manifestation of the condition. Coupled with this is the inherent variability in the intervention (physiotherapy), including the component treatments, duration and dose received by each patient.

Mechanisms to overcome this include to closely align the nature of the sample with the predictive variables explored, as with the study by Ogon et al. (2009) where patients were recruited with calcific tendinopathy and the majority of potential prognostic variables related to calcific deposits. An alternative approach is to substantially increase the sample size as a mechanism to increase the power of the study, as in the samples of 5,000 (Deutscher et al. 2009) and 1,030 (Chester et al. 2016). A different alternative is the use of a self-management programme based around a single exercise, with data from Littlewood et al. (2016) providing evidence in support of using a single

exercise which in turn would arguably address some of the issues around 'background noise' in the study.

Regardless of the approach taken, any potential prognostic variable must have acceptable levels of repeatability. Tools with low reliability (such as the SDT and ultrasound diagnoses in the current study) themselves introduce additional 'background noise', in parallel to limiting the confidence with which statistical conclusions can be viewed. Mechanisms by which the repeatability of any such tools can be improved is therefore a requisite stage for their use in future prognostic studies.

Given the wide-ranging nature of the physiotherapy intervention then it could be argued that there is merit in attempting to more closely define or constrain the content and dose of physiotherapy and/or the clinical pathway in which it sits. Whilst these approaches could be criticised for having limited clinical transferability, it could be argued that shaping the clinical environment in this manner could improve consistency in care as well as allowing more direct and constructive informing of future care by more relevant prognostic research.

One mechanism by which the above factors could be addressed is by designing clinical care with routine data collection embedded within it. This could occur in a defined component of the care pathway such as secondary care, whereby all patients who are referred in are routinely offered the opportunity to have routine data collection performed. This would allow for a much larger sample size with a potential reduction in potential biasing of those recruited. By analysing the baseline referral in, treatment and referral on patterns, then an informed mechanism for stratifying both care pathway and treatment could be undertaken.

As a next stage, prognostic factors at well-defined parts of the care pathway and/or where treatment is more narrowly defined could be undertaken. Such prognostic research would allow for both direct exploration of the utility of the prognostic models in a new sample (a key part of the prognostic framework; Kent et al. 2010) and also the subsequent refining of the care pathway whereby prognostic factors could be used to select specific patient cohorts for treatment. This cyclical mechanism would allow for maximal utility of the prognostic findings leading to a continual refinement of patient care alongside the opportunity for large scale, longitudinal cohort research.

7.10.3 Clinical implications

Regarding the potential prognostic factor of pain as measured by VAS, the study provided evidence that although some individuals in the cohort demonstrated substantial variation across their 3 VAS readings, the overall cohort VAS data demonstrated statistically significant correlations between the 3 measures taken, namely 'On activity', 'At rest' and 'At night'. As such these pain measures can be viewed as being inter-related, thereby providing insight into commonalities between the manifestation of pain in SIS/RCTendinopathy across different circumstances.

The comparatively low incidence of psychological symptoms at a level requiring monitoring or potential treatment in the SIS/RCTendinopathy cohort could be interpreted as evidence that clinicians need not be overly vigilant for indicators of poor mental health when managing such patients. However whilst the current study did not find psychological symptoms as being predictive this does not preclude the potential that sub-clinical levels of psychological symptoms might be present which the multi-factorial nature of physiotherapy intervention might be able to address.

The treatment measured in the current study provides a descriptive representation of physiotherapy delivered in an NHS setting for patients with SIS/RCTendinopathy. In highlighting the highly heterogeneous nature of the intervention, it could provide a foundation for stimulating debate as to whether the inherent variation is reflective of personalising treatment to the patient or simply inherent variability in the approaches of different clinicians. This should be balanced with the low predictive percentage from the current study as support for the opinion that SIS/RCTendinopathy is itself a highly complex, multi-factorial condition. This leads onto the main prognostic study findings which are that predicting outcome in patients with SIS/RCTendinopathy treated with physiotherapy is very difficult.

7.10.3.1 Prognosis

The clinical implications of this study are that there is value in clinicians considering baseline function and age of subjects for informing the likely change in patient reported pain and disability across the period of treatment and up to 3 months post discharge. Specifically those with greater impairment of function and disability at baseline are more likely to demonstrate greater improvement both at the point of discharge and for 3 months afterwards. Conversely the older the subject, the less likely they are to improve at the point of discharge. This can provide a context for discussions with patients regarding likely outcome and associated timescales. It can also provide a mechanism for patient triage where resources are limited.

However the findings also provide evidence that the ability to predict outcome in SIS/RCTendinopathy patients is limited and that where definitive answers are sought regarding absolute likelihood of improvement, current evidence is not able to guide this. One speculative explanation is that outcome may be independent of physiotherapy and at least partially reflect the natural course of the condition. If recovery is partially down to the natural course of the condition, then identifying the key ingredients of non-invasive, multimodal physiotherapy that optimise such recovery would be a noteworthy mechanism for optimising the use of limited clinical resources.

The identification of subjects that did not improve with physiotherapy reinforces the concept that some patients will not respond to this treatment approach. Identifying the nature of these patients is an equally important aspect of prognostic research so that such patients can be directed to other services or counselled on the limited therapeutic benefit of being referred for physiotherapy.

From the current study findings, the use of routine diagnostic ultrasound imaging to inform prognosis cannot be advocated. This provides partial triangulation with published guidelines recommending the use of diagnostic ultrasound if a first period of conservative management fails (Diercks et al. 2014). However multiple patients informally told MS that one of the reasons for them participating in the study was so that they could receive a scan. This highlights the issue of patient beliefs and the importance of managing expectations around imaging and care pathway. Nonetheless the potential for imaging to provide a mechanism for patients to derive a therapeutic benefit is worthy of future investigation.

7.11 Limitations

Numerous limitations must be acknowledged when considering the findings of this study. Firstly, the study was modest in size and undertaken by a single researcher. Consequently the timescale over which the study was undertaken and the consequent sample size were both limited. In light of the likely inherent ‘background noise’ within the study, the complex multi-factorial nature of the condition, poorly defined clinical care pathways and the highly heterogeneous nature of the intervention, the limited sample size and study duration limit the ability to accommodate such variation.

Due to the absence of a control group, any predictive capacity cannot be assigned to the intervention. Any improvements seen could be due instead to the natural course of the condition or interaction between the two.

Other categories that might be predictive of outcome such as patient or clinician expectation were not included in the study. Similarly it is possible that the particular variables and/or measurement tools used were not the ones likely to provide a predictive capacity. Therefore, whilst the majority of variables in the study were not found to be predictive, this does not mean that other aspects of these variable types would not be.

There is evidence in the literature that increasing experience as a Sonographer equates to improved consistency (Jeyam et al. 2008). As such it is possible that MS was improving in experience during the cohort data collection phase, thereby introducing additional inherent variability in diagnostic accuracy within the data. As such the use of a more established operator is advocated. Furthermore the specific shoulder pathology of patients comprising the ultrasound reproducibility study data were not recorded and so comparability with the cohort sample is limited.

In recording the treatment delivered it must be acknowledged that what was recorded was the treatment issued by the physiotherapist and not necessarily that received by the patient, particularly with regards to HEP where levels of adherence were not recorded.

7.12 Summary of discussion chapter

This study investigated factors across a wide range of categories as potential baseline predictors of change in functional status in patients with SIS/RCTendinopathy treated with physiotherapy in an NHS setting. Following a process of data reduction, seven variables were entered into a multivariate regression analysis for two different models, namely change in function from baseline to discharge and from baseline to 3 months post-discharge.

In the first model, total SPADI and age combined to explain 15.7% of the variance in OSS change between baseline and discharge and only total SPADI was retained in the second model, explaining 9.6% of the variance in OSS between baseline and 3 months post-discharge. Demographic (age), clinical history (duration of symptoms), patient reported measures (psychological symptoms), clinical measures (strength and scapular movement and control) and structural pathology via ultrasound imaging were not predictive in multi-variate analysis at any time point.

The low number of predictor variables and small amount of overall variation that was predicted by the models may be explained by the highly variable nature of (i) the pragmatic intervention and (ii) time from baseline to discharge. However the complex, multi-faceted nature of SIS/RCTendinopathy and the subjective experience of shoulder pain and disability highlight the challenge of predicting outcome, particularly where a relatively small sample size is used. Furthermore, the inclusion of two novel variables (SDT and diagnostic ultrasound) both with relatively poor reproducibility may also explain the low level of prediction.

The study findings provide some preliminary recommendations for future work in this area. These include an emphasis on including patient reported measures of function and disability in preference to physical measures. Alongside the likely predictive impact, parallel benefits of this include the reduced time and patient burden of testing thereby allowing for potentially greater sample sizes. Recognising the importance of statistical power and the threat of over-fitting, larger sample size and a focus on robustly measured variables is paramount.

In addition, the segmenting of groups of patients whose characteristics mean they are likely to receive particular treatment approaches may allow for a pragmatic reduction in variability within such studies. Furthermore, consideration of the length of time that patients are under treatment may also provide a similar focusing of future studies.

8 CONCLUSION

Currently available evidence to inform the prognosis of patients with SIS/RCTendinopathy who are treated with physiotherapy is of variable quality and provides limited consensus. However prognostic research has the potential to provide valuable information to inform patient expectation and guide clinicians with patient triage and counselling.

This study sought to contribute to the evidence base regarding predicting outcome in this patient group by undertaking analysis of a broad range of baseline variables collected in the clinical setting. These included the novel use of the scapular dyskinesis test and diagnostic ultrasound. A sequential process of data reduction was performed in order to tailor the number of candidate variables to the number of subjects recruited, so as to avoid the threat of overfitting.

The subsequent, most parsimonious model comprised baseline function and disability as measured by total SPADI and age as predictors of the change in the OSS between baseline and discharge. Specifically greater impairment of function and disability at baseline was predictive of greater improvement, whilst greater age was predictive of less improvement. For the change in the OSS between baseline and 3 months post-discharge, only baseline function and disability as measured by the total SPADI was predictive and in the same direction as for the baseline to discharge model. The percentage variance in the dependent variable explained by each model was low at 15.7% and 9.6% for the discharge and 3 months post-discharge models; respectively.

In light of the complex, multi-factorial nature of SIS/RCTendinopathy, the use of a pragmatic approach to treatment content and duration was a source of additional heterogeneity in the overall study. Furthermore the low levels of reproducibility of the 2 novel variables (scapular dyskinesis test and diagnostic ultrasound) retained for consideration in the prognostic models reduced their capacity to act as meaningful prognostic factors. These elements inevitably combined with the smaller than planned sample size to bring a reduction in overall statistical power to the study.

Along with the likely difficulty of predicting outcome in this patient group, the above methodological factors mean that the low predictive value of the models in this study are perhaps unsurprising. Mechanisms to optimise predictive ability in future studies include the use of prognostic research embedded within routine clinical care, so that minimally biased, large samples can be recruited. Development of mechanisms to optimise the reproducibility of any measurement tools used is also advocated.

As part of this, a more meaningful condensing of the current care pathways would provide a mechanism to address some additional sources of variability and also allow for more meaningful implementation of subsequent findings. Furthermore the quantifying and perhaps streamlining of the treatment 'ingredients' of physiotherapy would serve to similarly improve the research and subsequent clinical implementation environments.

Nonetheless, it is recommended that clinicians are made aware that baseline function and disability along with age may serve to partially predict outcome and that awareness of these aspects when triaging, assessing and treating patients will enable clinicians to tailor the management and communication with such patients, accordingly.

9 REFERENCE LIST

Adejumo, A. O. 2005. Effect of missing values on the Cohen's Kappa statistic for raters agreement measurement. *Int J Pure App Math* 22(1), pp. 13-31.

Ainsworth, R. et al. 2009. A prospective randomised placebo controlled clinical trial of a rehabilitation programme for patients with a diagnosis of massive rotator cuff tears of the shoulder. *Shoulder and Elbow* 1(1), pp. 55-60.

Alqunaee, M. et al. 2012. Diagnostic accuracy of clinical tests for subacromial impingement syndrome: a systematic review and meta-analysis. *Arch Phys Med Rehabil* 93(2), pp. 229-236.

Altman, D. 1991. *Practical statistics for medical research*. London: Chapman and Hall.

Altman, D. G. 2001. Systematic reviews of evaluations of prognostic variables. *BMJ* 323(7306), pp. 224-228.

Altman, D. G. and Royston, P. 2006. The cost of dichotomising continuous variables. *BMJ* 332(7549), p. 1080.

Altman, D. G. et al. 2009. Prognosis and prognostic research: validating a prognostic model. *BMJ* 338, p. b605.

Amstutz, H. C. et al. 1981. UCLA anatomic total shoulder arthroplasty. *Clin Orthop Relat Res* (155), pp. 7-20.

Ando, A. et al. 2013. Identification of prognostic factors for the nonoperative treatment of stiff shoulder. *Int Orthop* 37(5), pp. 859-864.

Angst, F. et al. 2004. Comprehensive assessment of clinical outcome and quality of life after total shoulder arthroplasty: usefulness and validity of subjective outcome measures. *Arthritis Rheum* 51(5), pp. 819-828.

Armstrong, A. 2014. Evaluation and management of adult shoulder pain: a focus on rotator cuff disorders, acromioclavicular joint arthritis, and glenohumeral arthritis. *Med Clin North Am* 98(4), pp. 755-775, xii.

Atalar, H. et al. 2009. Restricted scapular mobility during arm abduction: implications for impingement syndrome. *Acta Orthop Belg* 75(1), pp. 19-24.

Awatani, T. et al. 2016. Same-session and between-day intra-rater reliability of hand-held dynamometer measurements of isometric shoulder extensor strength. *J Phys Ther Sci* 28(3), pp. 936-939.

Aydogan, A. et al. 2003. Factors affecting therapeutic response of adhesive capsulitis in type II diabetes mellitus. *J Back Musc Rehab* 17, pp. 3-7.

Babyak, M. A. 2004. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med* 66(3), pp. 411-421.

Baertschi, E. et al. 2013. Interrater reliability of clinical tests to evaluate scapulothoracic motion. *BMC Musculoskelet Disord* 14, p. 315.

Bair, M. J. et al. 2003. Depression and pain comorbidity: a literature review. *Arch Intern Med* 163(20), pp. 2433-2445.

Balke, M. et al. 2013. Correlation of acromial morphology with impingement syndrome and rotator cuff tears. *Acta Orthop* 84(2), pp. 178-183.

Baring, T. et al. 2007. Management of rotator cuff disease: specific treatment for specific disorders. *B Prac Res Clin Rheum* 21(2), pp. 279-294.

Bartolozzi, A. et al. 1994. Determinants of outcome in the treatment of rotator cuff disease. *Clin Orthop Relat Res* (308), pp. 90-97.

Beaton, D. and Richards, R. R. 1998. Assessing the reliability and responsiveness of 5 shoulder questionnaires. *J Shoulder Elbow Surg* 7(6), pp. 565-572.

Beaton, D. E. et al. 2001. Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. *J Hand Ther* 14(2), pp. 128-146.

Beggs, I. 2011. Shoulder ultrasound. *Semin Ultrasound CT MR* 32(2), pp. 101-113.

Beggs, I. et al. 2010. *Musculoskeletal Ultrasound Technical Guidelines I. Shoulder*. Available at: http://www.essr.org/html/img/pool/shoulder.pdf [Accessed: 3/7/2014]

Bekkering, G. E. et al. 2005. Prognostic factors for low back pain in patients referred for physiotherapy: comparing outcomes and varying modeling techniques. *Spine* 30(16), pp. 1881-1886.

Bennell, K. et al. 2010. Efficacy of standardised manual therapy and home exercise programme for chronic rotator cuff disease: randomised placebo controlled trial. *BMJ* (7763), p. 82.

Binder, A. I. et al. 1984. Frozen shoulder: a long-term prospective study. *Ann Rheum Dis* 43(3), pp. 361-364.

Bjorck-van Dijken, C. et al. 2008. Low back pain, lifestyle factors and physical activity: a population based-study. *J Rehabil Med* 40(10), pp. 864-869.

Borstad, J. D. 2006. Resting position variables at the shoulder: evidence to support a posture-impairment association. *Phys Ther* 86(4), pp. 549-557.

Borstad, J. D. and Ludewig, P. M. 2002. Comparison of scapular kinematics between elevation and lowering of the arm in the scapular plane. *Clin Biomech* 17(9-10), pp. 650-659.

Bot, S. D. et al. 2005. Predictors of outcome in neck and shoulder symptoms: a cohort study in general practice. *Spine* 30(16), pp. E459-470.

Bouvier, B. et al. 2015. Upper Limb Kinematics Using Inertial and Magnetic Sensors: Comparison of Sensor-to-Segment Calibrations. Sensors (Basel) 15(8), pp. 18813-18833.

Bowerman, B. and O'Connell, R. 2000. *Linear Statistical Models: An Applied Approach*. Duxbury Classic.

Braun, C. et al. 2016. Prognostic Models in Adults Undergoing Physical Therapy for Rotator Cuff Disorders: Systematic Review. Phys Ther 96(7), pp. 961-971.

Brekke, M. et al. 2002. Severity of musculoskeletal pain: relations to socioeconomic inequality. *Soc Sci Med* 54(2), pp. 221-228.

Brooks, R. 1996. EuroQol: the current state of play. *Health Policy* 37(1), pp. 53-72.

Broughton, G., 2nd et al. 2006. The basic science of wound healing. *Plast Reconstr Surg* 117(7 Suppl), pp. 12s-34s.

Butt, U. et al. 2015. Does arthroscopic subacromial decompression improve quality of life. *Ann R Coll Surg Engl* 97(3), pp. 221-223.

Cadogan, A. et al. 2012. Clinical predictors of a positive response to guided diagnostic block into the subacromial bursa. *J Rehabil Med* 44(10), pp. 877-884.

Calis, M. et al. 2000. Diagnostic values of clinical diagnostic tests in subacromial impingement syndrome. *Ann Rheum Dis* 59(1), pp. 44-47.

Cardiff & Vale University Health Board. 2015a. *Cardiff & Vale University Health Board - MSK Physiotherapy*. Available at: http://www.cardiffandvaleuhb.wales.nhs.uk/page/77984 [Accessed: 2/11/2015].

Carr, A. J. et al. 2015. Clinical effectiveness and cost-effectiveness of open and arthroscopic rotator cuff repair [the UK Rotator Cuff Surgery (UKUFF) randomised trial]. *Health Technol Assess* 19(80), pp. 1-218.

Cassou, B. et al. 2002. Chronic neck and shoulder pain, age, and working conditions: longitudinal results from a large random sample in France. *Occup Environ Med* 59(8), pp. 537-544.

Castoldi, F. et al. 2009. External rotation lag sign revisited: accuracy for diagnosis of full thickness supraspinatus tear. *J Shoulder Elbow Surg* 18(4), pp. 529-534.

Cecchi, F. et al. 2014. Predictors of response to exercise therapy for chronic low back pain: result of a prospective study with one year follow-up. *Eur J Phys Rehabil Med* 50(2), pp. 143-151.

Chard, M. D. et al. 1988. The long-term outcome of rotator cuff tendinitis - a review study. *Br J Rheumatol* 27(5), pp. 385-389.

Chester, R. et al. 2016. Psychological factors are associated with the outcome of physiotherapy for people with shoulder pain: a multicentre longitudinal cohort study. *Br J Sports Med*, Published online first doi: 10.1136/bjsports-2016-096084

Chester, R. et al. 2013. Predicting response to physiotherapy treatment for musculoskeletal shoulder pain: a systematic review. *BMC Musculoskelet Disord* 14, p. 203.

Chien, C. W. et al. 2013. Comparative responsiveness of verbal and numerical rating scales to measure pain intensity in patients with chronic pain. *J Pain* 14(12), pp. 1653-1662.

Cloke, D. J. et al. 2005. A comparison of functional, patient-based scores in subacromial impingement. *J Shoulder Elbow Surg* 14(4), pp. 380-384.

Cobb, E. M. et al. 2014. Public Interest in Medical Research Participation: Differences by Volunteer Status and Study Type. Clin Transl Sci.

Cohen, J. et al. 2013. *Applied multiple regression/correlation analysis for the behavioral sciences*. 3rd edition ed. Mahwah, NJ: Lawrence Erlbaum Associates.

Collins, N. J. et al. 2010. Predictors of short and long term outcome in patellofemoral pain syndrome: a prospective longitudinal study. *BMC Musculoskelet Disord* 11, p. 11.

Conable, K. M. and Rosner, A. L. 2011. A narrative review of manual muscle testing and implications for muscle testing research. *J Chiropr Med* 10(3), pp. 157-165.

Concato, J. et al. 1993. The risk of determining risk with multivariable models. *Ann Intern Med* 118(3), pp. 201-210.

Conroy, D. E. and Hayes, K. W. 1998. The effect of joint mobilization as a component of comprehensive treatment for primary shoulder impingement syndrome. *J Orthop Sports Phys Ther* 28(1), pp. 3-14.

Consortium for the Accreditation of Sonographic Education (CASE). Available at: http://www.case-uk.org/ [Accessed 3/7/2014].

Contreras, F. et al. 2013. Predictors of success of corticosteroid injection for the management of rotator cuff disease. *Hss J* 9(1), pp. 2-5.

Cook, C. et al. 2014. The addition of cervical unilateral posterior-anterior mobilisation in the treatment of patients with shoulder impingement syndrome: a randomised clinical trial. *Man Ther* 19(1), pp. 18-24.

Cook, J. L. and Purdam, C. R. 2009. Is tendon pathology a continuum? A pathology model to explain the clinical presentation of load-induced tendinopathy. *Br J Sports Med* 43(6), pp. 409-416.

Cook, K. F. et al. 2003. Development and psychometric evaluation of the Flexilevel Scale of Shoulder Function. *Med Care* 41(7), pp. 823-835.

Cools, A. M. et al. 2014. Measuring shoulder external and internal rotation strength and range of motion: comprehensive intra-rater and inter-rater reliability study of several testing protocols. *J Shoulder Elbow Surg* 23(10), pp. 1454-1461.

Corazza, A. et al. 2015. Dynamic high-resolution ultrasound of the shoulder: how we do it. *Eur J Radiol* 84(2), pp. 266-277.

Crane, P. K. et al. 2006. A 37-item shoulder functional status item pool had negligible differential item functioning. *J Clin Epidemiol* 59(5), pp. 478-484.

Croft, P. et al. 1996. The clinical course of shoulder pain: prospective cohort study in primary care. Primary Care Rheumatology Society Shoulder Study Group. *BMJ* 313(7057), pp. 601-602.

Cummins, C. A. et al. 2009. Impingement syndrome: temporal outcomes of nonoperative treatment. *J Shoulder Elbow Surg* 18(2), pp. 172-177.

Dawson, J. et al. 1996. Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg Br* 78(4), pp. 593-600.

Dawson, J. et al. 2001. The benefits of using patient-based methods of assessment. Medium-term results of an observational study of shoulder surgery. *J Bone Joint Surg Br* 83(6), pp. 877-882.

de Winter, A. F. et al. 2004. Inter-observer reproducibility of measurements of range of motion in patients with shoulder pain using a digital inclinometer. *BMC Musculoskelet Disord* 5, p. 18.

Dean, B. J. et al. 2013. Why does my shoulder hurt? A review of the neuroanatomical and biochemical basis of shoulder pain. Br J Sports Med 47(17), pp. 1095-1104.

den Boeft, M. et al. 2016. The association between medically unexplained physical symptoms and health care use over two years and the influence of depressive and anxiety disorders and personality traits: a longitudinal study. *BMC Health Serv Res* 16, p. 100.

Derogatis, L. R. et al. 1974. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. *Behav Sci* 19(1), pp. 1-15.

Desai, A. S. et al. 2010. Critical appraisal of subjective outcome measures used in the assessment of shoulder disability. *Ann R Coll Surg Engl* 92(1), pp. 9-13.

Desjardins-Charbonneau, A. et al. 2015. The efficacy of manual therapy for rotator cuff tendinopathy: a systematic review and meta-analysis. J Orthop Sports Phys Ther 45(5), pp. 330-350.

Deutscher, D. et al. 2008. Implementing an integrated electronic outcomes and electronic health record process to create a foundation for clinical practice improvement. *Phys Ther* 88(2), pp. 270-285.

Deutscher, D. et al. 2009. Associations between treatment processes, patient characteristics, and outcomes in outpatient physical therapy practice. *Arch Phys Med Rehabil* 90(8), pp. 1349-1363.

Diercks, R. et al. 2014. Guideline for diagnosis and treatment of subacromial pain syndrome: a multidisciplinary review by the Dutch Orthopaedic Association. *Acta Orthop* 85(3), pp. 314-322.

Digi-Pas UK 2016. Available at: http://www.digipas.co.uk/products/digital-level-module/dwl-180.php [Accessed: 02/11/2016].

Dinant, G. J. et al. 2007. The necessary shift from diagnostic to prognostic research. *BMC Fam Pract* 8, p. 53.

Ditre, J. W. et al. 2011. Pain, Nicotine, and Smoking: Research Findings and Mechanistic Considerations. *Psychol Bull* 137(6), pp. 1065-1093.

Dixon, J. et al. 1998. Assessment of the reproducibility of clinical coding in routinely collected hospital activity data: a study in two hospitals. *J Public Health Med* 20(1), pp. 63-69.

Dogu, B. et al. 2012. Blind or ultrasound-guided corticosteroid injections and short-term response in subacromial impingement syndrome: a randomized, double-blind, prospective study. *Am J Phys Med Rehabil* 91(8), pp. 658-665.

Dong, W. et al. 2015. Treatments for shoulder impingement syndrome: a PRISMA systematic review and network meta-analysis. *Medicine* 94(10), p. e510.

Dorrestijn, O. et al. 2011. Patients with shoulder complaints in general practice: Consumption of medical care. *Rheumatology* 50(2), pp. 389-395.

Downie, W. W. et al. 1978. Studies with pain rating scales. *Ann Rheum Dis* 37(4), pp. 378-381.

Dragesund, T. and Kvale, A. 2016. Study protocol for Norwegian Psychomotor Physiotherapy versus Cognitive Patient Education in combination with active individualized physiotherapy in patients with long-lasting musculoskeletal pain - a randomized controlled trial. *BMC Musculoskelet Disord* 17, p. 325.

Ekeberg, O. M. et al. 2010a. Clinical, socio-demographic and radiological predictors of short-term outcome in rotator cuff disease. *BMC Musculoskelet Disord* 11, p. 239.

Ekeberg, O. M. et al. 2010b. A questionnaire found disease-specific WORC index is not more responsive than SPADI and OSS in rotator cuff disease. *J Clin Epidemiol* 63(5), pp. 575-584.

Ekeberg, O. M. et al. 2008. Agreement, reliability and validity in 3 shoulder questionnaires in patients with rotator cuff disease. *BMC Musculoskelet Disord* 9(68),

Engebretsen, K. et al. 2010. Predictors of shoulder pain and disability index (SPADI) and work status after 1 year in patients with subacromial shoulder pain. *BMC Musculoskelet Disord* 11, p. 218.

Ettinger, L. et al. 2014. Subacromial Injection Results in Further Scapular Dyskinesis. *Orthop J Sports Med* 2(8)

Factor, D. and Dale, B. 2014. Current concepts of rotator cuff tendinopathy. *Int J Sports Phys Ther* 9(2), pp. 274-288.

Ferraz, M. B. et al. 1990. Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. *J Rheumatol* 17(8), pp. 1022-1024.

Ferreira-Valente, M. A. et al. 2011. Validity of four pain intensity rating scales. *Pain* 152(10), pp. 2399-2404.

Field, A. 2009. *Discovering statistics using SPSS*. Sage publications.

Fillingim, R. B. et al. 2009. Sex, gender, and pain: a review of recent clinical and experimental findings. *J Pain* 10(5), pp. 447-485.

Gartner, J. and Heyer, A. 1995. Calcific tendinitis of the shoulder. *Orthopade* 24(3), pp. 284-302.

Gartsman, G. M. et al. 1998. Self-assessment of general health status in patients with five common shoulder conditions. *J Shoulder Elbow Surg* 7(3), pp. 228-237.

George, S. Z. et al. 2008. Investigation of elevated fear-avoidance beliefs for patients with low back pain: a secondary analysis involving patients enrolled in physical therapy clinical trials. *J Orthop Sports Phys Ther* 38(2), pp. 50-58.

George, S. Z. et al. 2006. Sex differences in predictors of outcome in selected physical therapy interventions for acute low back pain. *J Orthop Sports Phys Ther* 36(6), pp. 354-363.

Ginn, K. A. and Cohen, M. L. 2004. Conservative treatment for shoulder pain: prognostic indicators of outcome. *Arch Phys Med Rehabil* 85(8), pp. 1231-1235.

Girish, G. et al. 2011. Ultrasound of the shoulder: asymptomatic findings in men. *AJR Am J Roentgenol* 197(4), pp. W713-719.

Godfrey, J. et al. 2007. Reliability, validity, and responsiveness of the simple shoulder test: psychometric properties by age and injury type. *J Shoulder Elbow Surg* 16(3), pp. 260-267.

Godin, G. and Shephard, R. J. 1985. A simple method to assess exercise behavior in the community. *Can J Appl Sport Sci* 10(3), pp. 141-146.

Gorski, J. M. and Schwartz, L. H. 2003. Shoulder impingement presenting as neck pain. *J Bone Joint Surg Am* 85-a(4), pp. 635-638.

Gosens, T. and Hofstee, D. J. 2009. Calcifying tendinitis of the shoulder: advances in imaging and management. *Curr Rheum Rep* 11(2), pp. 129-134.

Green, S. et al. 2003. Physiotherapy interventions for shoulder pain. *Cochrane Database Syst Rev* (2):CD004258.

Gregg, C. D. et al. 2014. Prognostic factors associated with low back pain outcomes. *J Prim Health Care* 6(1), pp. 23-30.

Greving, K. et al. 2012. Incidence, prevalence, and consultation rates of shoulder complaints in general practice. *Scand J Rheumatol* 41(2), pp. 150-155.

Griggs, S. M. et al. 2000. Idiopathic adhesive capsulitis. A prospective functional outcome study of nonoperative treatment. *J Bone Joint Surg Am* 82-a(10)

Gruson, K. I. et al. 2008. Subacromial corticosteroid injections. *J Shoulder Elbow Surg* 17(1 Suppl), pp. 118s-130s.

Haghighat, S. et al. 2016. Effectiveness of Blind & Ultrasound Guided Corticosteroid Injection in Impingement Syndrome. *Glob J Health Sci* 8(7), pp. 179-184.

Hanchard, N. C. et al. 2013. Physical tests for shoulder impingements and local lesions of bursa, tendon or labrum that may accompany impingement. *Cochrane Database Syst Rev* (4), p. Cd007427.

Harniman, E. et al. 2004. Extracorporeal shock wave therapy for calcific and noncalcific tendonitis of the rotator cuff: A systematic review. *J Hand Ther* 17(2), pp. 132-151.

Harrington, S. et al. 2013. Upper extremity strength and range of motion and their relationship to function in breast cancer survivors. *Physiother Theory Pract* 29(7), pp. 513-520.

Hart, D. L. et al. 2006. Simulated computerized adaptive test for patients with shoulder impairments was efficient and produced valid measures of function. *J Clin Epidemiol* 59(3), pp. 290-298.

Hart, D. L. et al. 2010. A computerized adaptive test for patients with shoulder impairments produced responsive measures of function. *Phys Ther* 90(6), pp. 928-938.

Hart, D. L. et al. 2009. Screening for elevated levels of fear-avoidance beliefs regarding work or physical activities in people receiving outpatient therapy. *Phys Ther* 89(8), pp. 770-785.

Hawker, G. A. et al. 2011. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis Care Res* 63 Suppl 11, pp. S240-252.

Hayden, J. A. et al. 2006. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med* 144(6), pp. 427-437.

Hayes, K. et al. 2002. Shoulder instability: management and rehabilitation. *J Orthop Sports Phys Ther* 32(10), pp. 497-509.

Heald, S. L. et al. 1997. The shoulder pain and disability index: the construct validity and responsiveness of a region-specific disability measure. *Phys Ther* 77(10), pp. 1079-1089.

Hegedus, E. J. et al. 2008. Physical examination tests of the shoulder: a systematic review with meta-analysis of individual tests. *Br J Sports Med* 42(2), pp. 80-92

Hegedus, E. J. et al. 2012. Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests. *Br J Sports Med* 46(14), pp. 964-978.

Hemingway, H. et al. 2013. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ* 346, p. e5595.

Hemingway, H. et al. 2009. Ten steps towards improving prognosis research. *BMJ* 339, p. b4184.

Henkus, H. E. et al. 2009. Bursectomy compared with acromioplasty in the management of subacromial impingement syndrome: a prospective randomised study. *J Bone Joint Surg Br* 91(4), pp. 504-510.

Hill, J. C. and Fritz, J. M. 2011. Psychosocial influences on low back pain, disability, and response to treatment. *Phys Ther* 91(5), pp. 712-721.

Hingorani, A. D. et al. 2013. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ* 346, p. e5793.

Hjermstad, M. J. et al. 2011. Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: a systematic literature review. *J Pain Symptom Manage* 41(6), pp. 1073-1093.

Hoggan Health 2016. Available at: http://www.hogganhealth.net/pdfs/microfet2.pdf [Accessed: 28/10/2016].

Holmgren, T. et al. 2012. Effect of specific exercise strategy on need for surgery in patients with subacromial impingement syndrome: randomised controlled study. *BMJ* 344, p. e787.

Holtby, R. and Razmjou, H. 2004. Validity of the supraspinatus test as a single clinical test in diagnosing patients with rotator cuff pathology. *J Ortho Sports Phys Ther* 34(4), pp. 194-200.

Holtermann, A. et al. 2010. Prognostic factors for long-term sickness absence among employees with neck-shoulder and low-back pain. *Scand J Work Environ Health* 36(1), pp. 34-41.

Horn, M. E. et al. 2014. Fear of severe pain mediates sex differences in pain sensitivity responses to thermal stimuli. *Pain Res Treat* 2014, p. 897953.

Hsu, Y. H. et al. 2009. The effects of taping on scapular kinematics and muscle performance in baseball players with shoulder impingement syndrome. *J Electromyogr Kinesiol* 19(6), pp. 1092-1099.

http://www.strobe-statement.org. ; accessed 25th January 2018

Huber, W. et al. 2004. The German version of the Oxford Shoulder Score--cross-cultural adaptation and validation. *Arch Orthop Trauma Surg* 124(8), pp. 531-536.

Hughes, P. 2011. The Neer sign and Hawkins-Kennedy test for shoulder impingement. *J Physiother* 57(4), p. 260.

Hung, C. J. et al. 2010. Scapular kinematics and impairment features for classifying patients with subacromial impingement syndrome. *Man Ther* 15(6), pp. 547-551.

Hurst, N. P. et al. 1997. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *Br J Rheumatol* 36(5), pp. 551-559.

Iagnocco, A. et al. 2003. Sonographic study of painful shoulder. *Clin Exp Rheumatol* 21(3), pp. 355-358.

Ingwersen, K. G. et al. 2016. Ultrasound assessment for grading structural tendon changes in supraspinatus tendinopathy: an inter-rater reliability study. *BMJ Open* 6(5), p. e011746.

J Tech Medical 2016. Available at: http://www.jtechmedical.com/phocadownload/manuals/MN084-CommanderMuscleTesterManual.pdf [Accessed: 31/10/2016].

Jacobson, J. A. 2011. Shoulder US: anatomy, technique, and scanning pitfalls. *Radiology* 260(1), pp. 6-16.

Jaeschke, R. et al. 1989. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 10(4), pp. 407-415.

Jellema, P. et al. 2005. Why is a treatment aimed at psychosocial factors not effective in patients with (sub)acute low back pain? *Pain* 118(3), pp. 350-359.

Jensen, M. P. et al. 1986. The measurement of clinical pain intensity: a comparison of six methods. *Pain* 27(1), pp. 117-126.

Jeyam, M. et al. 2008. Are shoulder surgeons any good at diagnosing rotator cuff tears using ultrasound?: A comparative analysis of surgeon vs radiologist. *Int J Shoulder Surg* 2(1), pp. 4-6.

Jia, X. et al. 2009. Examination of the shoulder: the past, the present, and the future. *J Bone Joint Surg Am* 91 Suppl 6, pp. 10-18.

Johansson, K. et al. 2002. A combination of systematic review and clinicians' beliefs in interventions for subacromial pain. *Br J Gen Pract* 52(475), pp. 145-152.

Joseph, M. F. and Denegar, C. R. 2015. Treating tendinopathy: perspective on anti-inflammatory intervention and therapeutic exercise. *Clin Sports Med* 34(2), pp. 363-374.

Joyce, C. R. et al. 1975. Comparison of fixed interval and visual analogue scales for rating chronic pain. *Eur J Clin Pharmacol* 8(6), pp. 415-420.

Kaergaard, A. and Andersen, J. H. 2000. Musculoskeletal disorders of the neck and shoulders in female sewing machine operators: prevalence, incidence, and prognosis. *Occup Environ Med* 57(8), pp. 528-534.

Kalter, J. et al. 2011. Taping patients with clinical signs of subacromial impingement syndrome: The design of a randomized controlled trial. *BMC Musc Disord* 12(188)

Karels, C. H. et al. 2007. Social and psychological factors influenced the course of arm, neck and shoulder complaints. *J Clin Epidemiol* 60(8), pp. 839-848.

Katz, J. N. 2006. Lumbar disc disorders and low-back pain: socioeconomic factors and consequences. *J Bone Joint Surg Am* 88 Suppl 2, pp. 21-24.

Kelly, S. M. et al. 2010. The value of physical tests for subacromial impingement syndrome: a study of diagnostic accuracy. *Clin Rehabil* 24(2), pp. 149-158.

Kendall, F. P. et al. 1993. *Muscles: Testing and Function*. 4th ed. ed. Baltimore, MD: Williams & Wilkins.

Kennedy, C. A. et al. 2006. Prognosis in soft tissue disorders of the shoulder: predicting both change in disability and level of disability after treatment. *Phys Ther* 86(7), pp. 1013-1032

Kent, P. et al. 2010. Research methods for subgrouping low back pain. *BMC Med Res Methodol* 10, p. 62.

Kessel, L. and Watson, M. 1977. The painful arc syndrome. Clinical classification as a guide to management. *J Bone Joint Surg Br* 59(2), pp. 166-172.

Khan, Y. et al. 2013. The Painful Shoulder: Shoulder Impingement Syndrome. *Open Orthop J* 7, pp. 347-351.

Kibler, W. B. 1998. The role of the scapula in athletic shoulder function. *Am J Sports Med* 26(2), pp. 325-337.

Kibler, W. B. et al. 2002. Qualitative clinical evaluation of scapular dysfunction: a reliability study. *J Shoulder Elbow Surg* 11(6), pp. 550-556.

Kim, S. H. et al. 2004. Painful jerk test: a predictor of success in nonoperative treatment of posteroinferior instability of the shoulder. *Am J Sports Med* 32(8), pp. 1849-1855.

Kindler, L. L. et al. 2011. Sex differences in experimental and clinical pain sensitivity for patients with shoulder pain. *Eur J Pain* 15(2), pp. 118-123.

Koester, M. C. et al. 2005. Shoulder impingement syndrome. *Am J Med* 118(5), pp. 452-455.

Kohli, E. et al. 2009. Variability in the Hawthorne effect with regard to hand hygiene performance in high- and low-performing inpatient care units. *Infect Control Hosp Epidemiol* 30(3), pp. 222-225.

Kolber, M. J. et al. 2012. The reliability and concurrent validity of scapular plane shoulder elevation measurements using a digital inclinometer and goniometer. *Physiother Theory Pract* 28(2), pp. 161-168.

Kolber, M. J. et al. 2011. The reliability and minimal detectable change of shoulder mobility measurements using a digital inclinometer. *Physiother Theory Pract* 27(2), pp. 176-184.

Kooijman, M. et al. 2013. Patients with shoulder syndromes in general and physiotherapy practice: an observational study. *BMC Musculoskelet Disord* 14, p. 128.

Kooijman, M. K. et al. 2015. Pain intensity, neck pain and longer duration of complaints predict poorer outcome in patients with shoulder pain – a systematic review. *BMC Musc Disord* 16(1), p. 288.

Koorevaar, R. C. et al. 2016. Validation of the four-dimensional symptom questionnaire (4DSQ) and prevalence of psychological symptoms in orthopedic shoulder patients. *J Orthop Res* 34(4), pp. 683-691.

Koslow, P. A. et al. 2003. Specificity of the lateral scapular slide test in asymptomatic competitive athletes. *J Orthop Sports Phys Ther* 33(6), pp. 331-336.

Kottner, J. et al. 2011a. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol 64(1), pp. 96-106.

Kromer, T. O. et al. 2010. Effectiveness of individualized physiotherapy on pain and functioning compared to a standard exercise protocol in patients presenting with clinical signs of subacromial impingement syndrome. A randomized controlled trial. *BMC Musc Disord* 11(114)

Kromer, T. O. et al. 2013. Physiotherapy in patients with clinical signs of shoulder impingement syndrome: a randomized controlled trial. *J Rehabil Med* 45(5), pp. 488-497.

Kromer, T. O. et al. 2014. Influence of fear-avoidance beliefs on disability in patients with subacromial shoulder pain in primary care: a secondary analysis. *Phys Ther* 94(12), pp. 1775-1784.

Kuijpers, T. et al. 2006a. Clinical prediction rules for the prognosis of shoulder pain in general practice. *Pain* 120(3), pp. 276-285.

Kuijpers, T. et al. 2004. Systematic review of prognostic cohort studies on shoulder disorders. *Pain* 109(3), pp. 420-431.

Kuijpers, T. et al. 2006b. A prediction rule for shoulder pain related sick leave: a prospective cohort study. *BMC Musculoskelet Disord* 7, p. 97.

Kuroda, S. et al. 2001. The natural course of atraumatic shoulder instability. *J Shoulder Elbow Surg* 10(2), pp. 100-104.

Lakke, S. E. et al. 2009. Risk and prognostic factors for non-specific musculoskeletal pain: a synthesis of evidence from systematic reviews classified into ICF dimensions. *Pain* 147(1-3), pp. 153-164.

Lalkhen, A. G. and McCluskey, A. 2008. Clinical tests: sensitivity and specificity. *Contin Educ Anaesth Crit Care Pain* 8 (6): 221-223

Landis, J. R. and Koch, G. G. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33(1), pp. 159-174.

Laslett, M. et al. 2015. Shoulder pain in primary care--part 2: predictors of clinical outcome to 12 months. *J Rehabil Med* 47(1), pp. 66-71.

Lewis, J. S. 2009. Rotator cuff tendinopathy/subacromial impingement syndrome: is it time for a new method of assessment? *Br J Sports Med* 43(4), pp. 259-264

Lewis, J. S. 2010. Rotator cuff tendinopathy: a model for the continuum of pathology and related management. *Br J Sports Med* 44(13), pp. 918-923.

Lewis, J. S. 2012. A specific exercise program for patients with subacromial impingement syndrome can improve function and reduce the need for surgery. *J Physiother* 58(2), p. 127.

Linaker, C. H. and Walker-Bone, K. 2015. Shoulder disorders and occupation. *Best Pract Res Clin Rheumatol* 29(3), pp. 405-423.

Littlewood, C. et al. 2013. Epidemiology of Rotator Cuff Tendinopathy: A Systematic Review. Shoulder & Elbow 5(4).

Littlewood, C. et al. 2015. Therapeutic exercise for rotator cuff tendinopathy: a systematic review of contextual factors and prescription parameters. Int J Rehabil Res 38(2), pp. 95-106.

Littlewood, C. et al. 2016. A self-managed single exercise programme versus usual physiotherapy treatment for rotator cuff tendinopathy: a randomised controlled trial (the SELF study). *Clin Rehabil* 30(7), pp. 686-696.

Lopes, A. D. et al. 2015. Visual scapular dyskinesis: kinematics and muscle activity alterations in patients with subacromial impingement syndrome. *Arch Phys Med Rehabil* 96(2), pp. 298-306.

Lorig, K. et al. 1989. Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. *Arthritis Rheum* 32(1), pp. 37-44.

Lu, T. W. et al. 2007. Enhancing the examiner's resisting force improves the reliability of manual muscle strength measurements: comparison of a new device with hand-held dynamometry. *J Rehabil Med* 39(9), pp. 679-684.

Ludewig, P. M. et al. 2013. What's in a name? Using movement system diagnoses versus pathoanatomic diagnoses. *J Orthop Sports Phys Ther* 43(5), pp. 280-283.

Luime, J. J. et al. 2004. Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol* 33(2), pp. 73-81.

Macfarlane, G. J. et al. 1998. Predictors of chronic shoulder pain: a population based prospective study. *J Rheumatol* 25(8), pp. 1612-1615.

Macznik, A. K. et al. 2015. Online technology use in physiotherapy teaching and learning: a systematic review of effectiveness and users' perceptions. *BMC Med Educ* 15, p. 160.

Mallows, A. et al. 2017. Association of psychological variables and outcome in tendinopathy: a systematic review. Br J Sports Med 51(9), pp. 743-748.

Mao, C. Y. et al. 1997. Frozen shoulder: correlation between the response to physical therapy and follow-up shoulder arthrography. *Arch Phys Med Rehabil* 78(8), pp. 857-859.

McClinton, S. M. et al. 2015. Predictors of response to physical therapy intervention for plantar heel pain. *Foot Ankle Int* 36(4), pp. 408-416.

McClure, P. et al. 2009. A Clinical Method for Identifying Scapular Dyskinesis, Part 1: Reliability. *J Athl Train* 44(2), pp. 160-164.

McClure, P. W. et al. 2006. Shoulder function and 3-dimensional scapular kinematics in people with and without shoulder impingement syndrome. *Phys Ther* 86(8), pp. 1075-1090.

McCreesh, K. M. et al. 2016. Ultrasound measures of supraspinatus tendon thickness and acromiohumeral distance in rotator cuff tendinopathy are reliable. *J Clin Ultrasound* 44(3), pp. 159-166.

McDonald, S. et al. 2010. Basic appearance of ultrasound structures and pitfalls. *Phys Med Rehabil Clin N Am* 21(3), pp. 461-479.

McHorney, C. A. et al. 1994. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 32(1), pp. 40-66.

McHugh, M. L. 2012. Interrater reliability: the kappa statistic. *Biochem Med* 22(3), pp. 276-282.

McNally, E. 2014. *Practical Musculoskeletal Ultrasound*. 2nd ed. Elsevier.

Messina, C. et al. 2016. Ultrasound-guided interventional procedures around the shoulder. *Br J Radiol* 89(1057), p. 20150372.

Michener, L. A. et al. 2003. Anatomical and biomechanical mechanisms of subacromial impingement syndrome. *Clin Biomech* 18(5), pp. 369-379.

Michener, L. A. et al. 2009. Reliability and Diagnostic Accuracy of 5 Physical Examination Tests and Combination of Tests for Subacromial Impingement. *Arch Phys Med Rehabil* 90(11), pp. 1898-1903.

Micheroli, R. et al. 2015. Correlation of findings in clinical and high resolution ultrasonography examinations of the painful shoulder. *J Ultrason* 15(60), pp. 29-44.

Miles, J. and Shevlin, M. 2001. *Applying Regression and Correlation: A Guide for Students and Researchers*. SAGE.

Min, K. S. et al. 2013. A double-blind randomized controlled trial comparing the effects of subacromial injection with corticosteroid versus NSAID in patients with shoulder impingement syndrome. *J Shoulder Elbow Surg* 22(5), pp. 595-601.

Minns Lowe, C. J. et al. 2014. Living with a symptomatic rotator cuff tear 'bad days, bad nights': a qualitative study. *BMC Musculoskelet Disord* 15, p. 228.

Mintken, P. E. et al. 2010. Some factors predict successful short-term outcomes in individuals with shoulder pain receiving cervicothoracic manipulation: a single-arm trial. *Phys Ther* 90(1), pp. 26-42.

Miranda, H. et al. 2001. A prospective study of work related factors and physical exercise as predictors of shoulder pain. *Occup Environ Med* 58(8), pp. 528-534.

Mitchell, C. et al. 2005. Shoulder pain: Diagnosis and management in primary care. *BMJ* 331(7525), pp. 1124-1128.

Mole, D. et al. 1993. Results of endoscopic treatment of non-broken tendinopathies of the rotator cuff. 2. Calcifications of the rotator cuff. *Rev Chir Orthop Reparatrice Appar Mot* 79(7), pp. 532-541.

Moons, K. G. et al. 2009a. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 338, p. b606.

Moons, K. G. et al. 2009b. Prognosis and prognostic research: what, why, and how? *BMJ* 338, p. b375.

Mori, B. et al. 2016. Canadian Physiotherapy Assessment of Clinical Performance: Face and Content Validity. *Physiother Can* 68(1), pp. 64-72.

Morrison, D. S. et al. 1997. Non-operative treatment of subacromial impingement syndrome. *J Bone Joint Surg Am* 79(5), pp. 732-737.

Mullaney, M. J. et al. 2010. Reliability of shoulder range of motion comparing a goniometer to a digital level. Physiother Theory Pract 26(5), pp. 327-333.

Myers, R. 2000. *Classical and Modern Regression with Applications*. Duxbury Classic.

Naredo, E. et al. 2002. Painful shoulder: comparison of physical examination and ultrasonographic findings. *Ann Rheum Dis* 61(2), pp. 132-136.

Ndlovu, M. et al. 2014. Pain medication management of musculoskeletal conditions at first presentation in primary care: analysis of routinely collected medical record data. *BMC Musculoskelet Disord* 15, p. 418.

Neer, C. S. 1972. Anterior acromioplasty for the chronic impingement syndrome in the shoulder: a preliminary report. *J Bone Joint Surg Am* 54(1), pp. 41-50.

Neer, C. S., 2nd. 1983. Impingement lesions. *Clin Orthop Relat Res* (173), pp. 70-77.

NHMRC. 2000. *How to use the evidence: assessment and application of scientific evidence*. Canberra: National Health and Medical Research Council. Available at: https://www.nhmrc.gov.au/guidelines-publications/cp69 [Accessed: 4/4/2016]

Nicholas, M. K. 2007. The pain self-efficacy questionnaire: Taking pain into account. *Eur J Pain* 11(2), pp. 153-163.

Ogon, P. et al. 2009. Prognostic factors in nonoperative therapy for chronic symptomatic calcific tendinitis of the shoulder. *Arthritis Rheum* 60(10), pp. 2978-2984.

Ostor, A. J. et al. 2005. Diagnosis and relation to general health of shoulder disorders presenting to primary care. *Rheumatology* 44(6), pp. 800-805.

Ottenheijm, R. P. et al. 2010. Accuracy of diagnostic ultrasound in patients with suspected subacromial disorders: A systematic review and meta-analysis. *Arch Phys Med Rehabil* 91(10), pp. 1616-1625.

Pandey, V. et al. 2016. Does scapular morphology affect the integrity of the rotator cuff? *J Shoulder Elbow Surg* 25(3), pp. 413-421.

Park, H. B. et al. 2005. Diagnostic accuracy of clinical tests for the different degrees of subacromial impingement syndrome. *J Bone Joint Surg Am* 87(7), pp. 1446-1455.

Parsons, I. M. t. et al. 2004. Glenohumeral arthritis and its management. Phys Med Rehabil Clin N Am 15(2), pp. 447-474. doi: 10.1016/j.pmr.2003.12.001

Payne, C. and Michener, L. A. 2014. Physiotherapists use of and perspectives on the importance of patient-reported outcome measures for shoulder dysfunction. *Shoulder Elbow* 6(3), pp. 204-214.

Peduzzi, P. et al. 1996. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 49(12), pp. 1373-1379.

Phillips, B. et al. 2016. Oxford Centre for Evidence-Based Medicine Levels of Evidence. Oxford: Centre for Evidence-Based Medicine. Available at: http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/ [Accessed: 4/11/2016].

Phyomaung, P. P. et al. 2014. Are depression, anxiety and poor mental health risk factors for knee pain? A systematic review. *BMC Musculoskelet Disord* 15, p. 10.

Pincus, T. et al. 2002. A systematic review of psychological factors as predictors of chronicity/disability in prospective cohorts of low back pain. *Spine* 27(5), pp. E109-120.

Portney, L. G. and Watkins, M. P. 2007. *Foundations of Clinical Research: Applications to Practice*. Prentice Hall, Lebanon, Indiana, U.S.A.

Rangan, A. et al. 2015. Surgical vs nonsurgical treatment of adults with displaced fractures of the proximal humerus: the PROFHER randomized clinical trial. *JAMA* 313(10), pp. 1037-1047.

Read, J. W. and Perko, M. 1998. Shoulder ultrasound: diagnostic accuracy for impingement syndrome, rotator cuff tear, and biceps tendon pathology. *J Shoulder Elbow Surg* 7(3), pp. 264-271.

Reilingh, M. L. et al. 2008. Course and prognosis of shoulder symptoms in general practice. *Rheumatology* 47(5), pp. 724-730.

Reinold, M. M. et al. 2004. Electromyographic analysis of the rotator cuff and deltoid musculature during common shoulder external rotation exercises. *J Orthop Sports Phys Ther* 34(7), pp. 385-394.

Rihn, J. A. et al. 2011. Duration of symptoms resulting from lumbar disc herniation: effect on treatment outcomes: analysis of the Spine Patient Outcomes Research Trial (SPORT). *J Bone Joint Surg Am* 93(20), pp. 1906-1914.

Riley, R. D. et al. 2013. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med* 10(2), p. e1001380.

Roach, K. E. et al. 1991. Development of a shoulder pain and disability index. *Arthritis Care Res* 4(4), pp. 143-149.

Roddey, T. S. et al. 2000. Comparison of the University of California-Los Angeles Shoulder Scale and the Simple Shoulder Test with the shoulder pain and disability index: single-administration reliability and validity. *Phys Ther* 80(8), pp. 759-768.

Roe, Y. et al. 2013. A systematic review of measures of shoulder pain and functioning using the International classification of functioning, disability and health (ICF). *BMC Musculoskelet Disord* 14, p. 73.

Roll, S. C. et al. 2016. Clinical utilization of musculoskeletal sonography involving non-physician rehabilitation providers: A scoping review. *Eur J Phys Rehabil Med* 52(2), pp. 253-262.

Royston, P. et al. 2009. Prognosis and prognostic research: Developing a prognostic model. *BMJ* 338, p. b604.

Ryall, C. et al. 2007. A prospective cohort study of arm pain in primary care and physiotherapy--prognostic determinants. *Rheumatology* 46(3), pp. 508-515.

Saltychev, M. et al. 2015. Conservative treatment or surgery for shoulder impingement: systematic review and meta-analysis. *Disabil Rehabil* 37(1), pp. 1-8.

Schellingerhout, J. M. et al. 2008. Lack of uniformity in diagnostic labeling of shoulder pain: time for a different approach. Man Ther 13(6), pp. 478-483.

Schmidt, S. et al. 2014. Evaluation of shoulder-specific patient-reported outcome measures: a systematic and standardized comparison of available evidence. *J Shoulder Elbow Surg* 23(3), pp. 434-444.

Schmitz, C. et al. 2015. Efficacy and safety of extracorporeal shock wave therapy for orthopedic conditions: a systematic review on studies listed in the PEDro database. *Br Med Bull* 116, pp. 115-138.

Schrama, P. P. et al. 2014. Intraexaminer reliability of hand-held dynamometry in the upper extremity: a systematic review. *Arch Phys Med Rehabil* 95(12), pp. 2444-2469.

Seitz, A. L. et al. 2011. Mechanisms of rotator cuff tendinopathy: intrinsic, extrinsic, or both? *Clin Biomech* 26(1), pp. 1-12.

Selfe, J. et al. 2016. Are there three main subgroups within the patellofemoral pain population? A detailed characterisation study of 127 patients to help develop targeted intervention (TIPPs). Br J Sports Med 50(14), pp. 873-880.

Senbursa, G. et al. 2007. Comparison of conservative treatment with and without manual physical therapy for patients with shoulder impingement syndrome: A prospective, randomized clinical trial. *Knee Surg Sport Traum Arthro* 15(7), pp. 915-921.

Shaffer, B. et al. 1992. Frozen shoulder. A long-term follow-up. *J Bone Joint Surg Am* 74(5), pp. 738-746.

Shrout, P. E. and Fleiss, J. L. 1979. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86(2), pp. 420-428.

Sim, J. and Wright, C. C. 2005. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 85(3), pp. 257-268.

Sindhu, B. S. et al. 2012. Influence of fear-avoidance beliefs on functional status outcomes for people with musculoskeletal conditions of the shoulder. *Phys Ther* 92(8), pp. 992-1005.

Smith, A. M. et al. 2005. Rotator cuff repair in patients with rheumatoid arthritis. *J Bone Joint Surg Am* 87(8), pp. 1782-1787.

Smith, M. et al. 2009. Upper and lower trapezius muscle activity in subjects with subacromial impingement symptoms: Is there imbalance and can taping change it? *Phys Ther Sport* 10(2), pp. 45-50.

Smith, M. J. et al. 2015. A training, assessment and feedback package for the trainee shoulder sonographer. *Ultrasound* 23(1), pp. 29-41.

Solomon, D. H. et al. 2001. Referrals for musculoskeletal disorders: patterns, predictors, and outcomes. *J Rheumatol* 28(9), pp. 2090-2095.

Soslowsky, L. J. et al. 2002. Rotator cuff tendinosis in an animal model: role of extrinsic and overuse factors. *Ann Biomed Eng* 30(8), pp. 1057-1063.

Spiegl, U. J. et al. 2014. Symptomatic internal impingement of the shoulder in overhead athletes. *Sports Med Arthrosc* 22(2), pp. 120-129.

Steffens, D. et al. 2014. Prognosis of chronic low back pain in patients presenting to a private community-based group exercise program. *Eur Spine J* 23(1), pp. 113-119.

Steuri, R. et al. 2017. Effectiveness of conservative interventions including exercise, manual therapy and medical management in adults with shoulder impingement: a systematic review and meta-analysis of RCTs. Br J Sports Med 51(18), pp. 1340-1347.

Steyerberg, E. W. et al. 2001. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. *Med Decis Making* 21(1), pp. 45-56.

Struyf, F. and Meeus, M. 2014. Current evidence on physical therapy in patients with adhesive capsulitis: what are we missing? *Clin Rheumatol* 33(5), pp. 593-600.

Sullivan, M. J. et al. 2008. Stage of chronicity and treatment response in patients with musculoskeletal injuries and concurrent symptoms of depression. *Pain* 135(1-2), pp. 151-159.

Sullivan, M. J. L. et al. 1995. The Pain Catastrophizing Scale: Development and validation. *Psych Assess* 7(4), p. 524.

Sun, G. W. et al. 1996. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol* 49(8), pp. 907-916.

Tanaka, K. et al. 2010. Joint mobilization versus self-exercises for limited glenohumeral joint mobility: randomized controlled study of management of rehabilitation. *Clin Rheumatol* 29(12), pp. 1439-1444.

Tashjian, R. Z. et al. 2009. Minimal clinically important differences (MCID) and patient acceptable symptomatic state (PASS) for visual analog scales (VAS) measuring pain in patients treated for rotator cuff disease. *J Shoulder Elbow Surg* 18(6), pp. 927-932.

Tate, A. et al. 2004. Validity of a visual classification system for scapular motion. *J Orthop Sport Phys Ther* 34(1), p. A42.

Tate, A. R. et al. 2009. A Clinical Method for Identifying Scapular Dyskinesis, Part 2: Validity. *J Athl Train* 44(2), pp. 165-173.

Tebbe, B. B. et al. 2013. Assessing psychological health in midwifery practice: a validation study of the Four-Dimensional Symptom Questionnaire (4DSQ), a Dutch primary care instrument. *Midwifery* 29(6), pp. 608-615.

Tempelaere, C. et al. 2016. Dynamic Three-Dimensional Shoulder Mri during Active Motion for Investigation of Rotator Cuff Diseases. *PLoS One* 11(7), p. e0158563.

Terluin, B. et al. 2014a. To what extent does the anxiety scale of the Four-Dimensional Symptom Questionnaire (4DSQ) detect specific types of anxiety disorder in primary care? A psychometric study. *BMC Psychiatry* 14, p. 121.

Terluin, B. et al. 2014b. The English version of the four-dimensional symptom questionnaire (4DSQ) measures the same as the original Dutch questionnaire: a validation study. *Eur J Gen Pract* 20(4), pp. 320-326.

Terluin, B. et al. 2006. The Four-Dimensional Symptom Questionnaire (4DSQ): a validation study of a multidimensional self-report questionnaire to assess distress, depression, anxiety and somatization. *BMC Psychiatry* 6, p. 34.

The Royal College of Radiologists. 2012. *Ultrasound training recommendations for medical and surgical specialities*, *Second edition*. London. Available at: https://www.rcr.ac.uk/sites/default/files/publication/BFCR(12)17_ultrasound_training.pdf [Accessed 4/12/13]

Thoomes-de Graaf, M. et al. 2014. Inter-professional agreement of ultrasound-based diagnoses in patients with shoulder pain between physical therapists and radiologists in the Netherlands. *Man Ther* 19(5), pp. 478-483.

Timmons, M. K. et al. 2012. Scapular kinematics and subacromial-impingement syndrome: a meta-analysis. *J Sport Rehabil* 21(4), pp. 354-370.

Turchin, D. C. et al. 1998. Validity of observer-based aggregate scoring systems as descriptors of elbow pain, function, and disability. *J Bone Joint Surg Am* 80(2), pp. 154-162.

Turman, K. A. et al. 2010. Massive Rotator Cuff Tear in an Adolescent Athlete: A Case Report. *Sports Health* 2(1), pp. 51-55.

Tyler, T. F. et al. 2010. Correction of posterior shoulder tightness is associated with symptom resolution in patients with internal impingement. *Am J Sports Med* 38(1), pp. 114-119.

UK Government Web Archive – The National Archives. Available at: http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/census/2011-census-analysis/local-area-analysis-of-qualifications-across-england-and-wales/rpt---local-area-analysis-of-qualifications-across-england-and-wales.html [Accessed: 17 02 2017].

Vahlensieck, M. 2000. MRI of the shoulder. *Eur Radiol* 10(2), pp. 242-249.

van der Heijden, G. J. 1999. Shoulder disorders: a state-of-the-art review. *Best Pract Res Clin Rheumatol* 13(2), pp. 287-309.

van der Linde, J. A. et al. 2015. The Oxford Shoulder Instability Score; validation in Dutch and first-time assessment of its smallest detectable change. *J Orthop Surg Res* 10, p. 146.

van der Windt, D. A. et al. 1996. Shoulder disorders in general practice: prognostic indicators of outcome. *Br J Gen Pract* 46(410), pp. 519-523.

van der Windt, D. A. et al. 1995. Shoulder disorders in general practice: incidence, patient characteristics, and management. *Ann Rheum Dis* 54(12), pp. 959-964.

van der Windt, D. A. et al. 2007. Do psychological factors predict outcome in both low-back pain and shoulder pain? *Ann Rheum Dis* 66(3), pp. 313-319.

van Eijsden-Besseling, M. D. et al. 2010. The influence of work and treatment related factors on clinical status and disability in patients with non-specific work-related upper limb disorders. *Work* 37(4), pp. 425-432.

van Rijn, R. M. et al. 2010. Associations between work-related factors and specific disorders of the shoulder--a systematic review of the literature. *Scand J Work Environ Health* 36(3), pp. 189-201.

Vargas-Prada, S. and Coggon, D. 2015. Psychological and psychosocial determinants of musculoskeletal pain and associated disability. *Best Pract Res Clin Rheumatol* 29(3), pp. 374-390.

Verbunt, J. A. et al. 2005. Pain-related factors contributing to muscle inhibition in patients with chronic low back pain: an experimental investigation based on superimposed electrical stimulation. *Clin J Pain* 21(3), pp. 232-240.

Verkerk, K. et al. 2015. Prognosis and course of pain in patients with chronic non-specific low back pain: A 1-year follow-up cohort study. *Eur J Pain* 19(8), pp. 1101-1110.

Vermeulen, H. M. et al. 2006. Comparison of high-grade and low-grade mobilization techniques in the management of adhesive capsulitis of the shoulder: randomized controlled trial. *Phys Ther* 86(3), pp. 355-368.

Viera, A. J. and Garrett, J. M. 2005. Understanding interobserver agreement: the kappa statistic. *Fam Med* 37(5), pp. 360-363.

Viester, L. et al. 2013. The relation between body mass index and musculoskeletal symptoms in the working population. *BMC Musculoskelet Disord* 14, p. 238.

Viikari-Juntura, E. et al. 2000. Predictive validity of symptoms and signs in the neck and shoulders. *J Clin Epidemiol* 53(8), pp. 800-808.

Viikari-Juntura, E. et al. 2008. Risk factors of atherosclerosis and shoulder pain--is there an association? A systematic review. Eur J Pain 12(4), pp. 412-426.

Virta, L. et al. 2009. How many patients with subacromial impingement syndrome recover with physiotherapy? A follow-up study of a supervised exercise programme. *Adv Physiother* 11(3), pp. 166-173.

Vlaeyen, J. W. and Linton, S. J. 2000. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *Pain* 85(3), pp. 317-332.

Vranceanu, A. M. et al. 2009. Psychosocial aspects of disabling musculoskeletal pain. *J Bone Joint Surg Am* 91(8), pp. 2014-2018.

Waddell, G. 2004. *The Back Pain Revolution*. Churchill Livingstone.

Waddell, G. et al. 1993. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain* 52(2), pp. 157-168.

Walker-Bone, K. E. et al. 2003. Criteria for assessing pain and nonarticular soft-tissue rheumatic disorders of the neck and upper limb. *Semin Arthritis Rheum* 33(3), pp. 168-184.

Wang, Y. C. et al. 2010. Translating shoulder computerized adaptive testing generated outcome measures into clinical practice. *J Hand Ther* 23(4), pp. 372-382; quiz 383.

Ware, J. E., Jr. et al. 1995. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. *Med Care* 33(4 Suppl), pp. As264-279.

Westrick, R. B. et al. 2013. Isometric shoulder strength reference values for physically active collegiate males and females. *Sports Health* 5(1), pp. 17-21.

Winokur, A. et al. 1984. Symptoms of emotional distress in a family planning service: stability over a four-week period. *Br J Psychiatry* 144, pp. 395-399.

Winters, J. C. et al. 1999. The long-term course of shoulder complaints: a prospective study in general practice. *Rheumatology* 38(2), pp. 160-163.

Woby, S. R. et al. 2005. Psychometric properties of the TSK-11: a shortened version of the Tampa Scale for Kinesiophobia. *Pain* 117(1-2), pp. 137-144.

World Health Organization 2016. International Classification of Functioning, Disability and Health (ICF). Available at: http://www.who.int/classifications/icf/en/ [Accessed 02/11/2016].

Yang, J. L. et al. 2008. Shoulder kinematic features using arm elevation and rotation tests for classifying patients with frozen shoulder syndrome who respond to physical therapy. *Man Ther* 13(6), pp. 544-551.

Yang, J. L. and Lin, J. J. 2006. Reliability of function-related tests in patients with shoulder pathologies. *J Orthop Sports Phys Ther* 36(8), pp. 572-576.

Younis, F. et al. 2011. The range of the Oxford Shoulder Score in the asymptomatic population: a marker for post-operative improvement. *Ann R Coll Surg Engl* 93(8), pp. 629-633.

Zheng, X. et al. 2005. Data from a study of effectiveness suggested potential prognostic factors related to the patterns of shoulder pain. *J Clin Epidemiol* 58(8), pp. 823-830.

Zuckerman, J. D. and Rokito, A. 2011. Frozen shoulder: a consensus definition. *J Shoulder Elbow Surg* 20(2), pp. 322-325.

10 Appendix II: Literature Review

Characteristics of the relevant studies, summarised in table form.

PROBAST process

PROBAST outcomes

PROBAST Overall judgement

10.1 Characteristics of the relevant studies, summarised in table form

Bartolozzi et al. (1994)

Study design	Retrospective cohort study
Pathology; inclusion/exclusion	'Rotator cuff disease' All pts had X-rays, 101 had MRI; for other 35 patients diagnosis was based entirely on history and clinical examination – comprising painful arc, +ve impingement sign, +ve impingement test, specific rotator cuff weakness. None of the 35 had a +ve drop arm test or clinically evident RC weakness
Patient care pathway and clinical setting	Patients attending orthopaedic department with impingement syndrome treated non-operatively
Sample size and nature of those not consenting	226 consecutive patients. 90 excluded (inadequate f/u (34), radiographic or clinical evidence of GH or ACJ OA (32), h/o neurologic disease (incl diabetes and cervical radiculopathy) (18), h/o frozen shoulder (4), h/o prox humeral # (2). Thus 136 patients were included. No information on those who did or did not consent
Potential prognostic factors and how they were measured (A-E)	• Demographics factors: Age, gender, occupation • Clinical history factors: Side-related symptoms: dominant arm involvement, Reason for symptom onset, Symptom duration • Patient reported measures: Pain (none), Psychological symptoms (none), Function / Disability (UCLA) • Clinical measures: Strength (shoulder girdle weakness), ROM (active abduction and flexion ROM), Scapular movement and control (none) • Structural pathology: Imaging (MRI: Rotator cuff pathology), Orthopaedic tests (none)
Treatment (type; duration and frequency)	Type: rotator cuff strengthening and ROM exercises
Outcome (how defined; timing; loss to follow up)	Outcome via combination of UCLA score at discharge and perceived improvement at average of 20 months Loss to follow up, failed to compare the baseline characteristics of those who were and those who were not followed up
Analysis (how candidate variables selected, incl no cases per variable) (Multivariate regression analysis approach)	No attempt made to trim the number of candidate variables Number of cases per variable = 10 R^2 values not presented

Chester et al. (2016)

Study design	Prospective cohort study
Pathology; inclusion/exclusion	Non-specific shoulder pain Patients were eligible if they were aged 18 years or older and described shoulder or arm pain aggravated by shoulder movements. Patients with significant reproduction of shoulder pain on spinal movement, or greater reproduction on spinal movement compared to shoulder movement, were excluded from the study. Patients with the following aetiology for shoulder pain were excluded: radiculopathy, postsurgery, postfracture, posttraumatic dislocation or systemic source.
Patient care pathway and clinical setting	Patients referred to physiotherapy for the management of musculoskeletal shoulder pain at 11 NHS trusts and social enterprises in the East of England were recruited Participating physiotherapy departments were located within primary and secondary care
Sample size and nature of those not consenting	Based on analysis with a general linear model, 780 participants provided 90% statistical power to detect an effect size of <0.25 SDs adjusted for other variables with a coefficient of determination with the outcome of up to 0.5. n=1030 Yes - the characteristics of those who did and did not consent should be considered, so as to avoid potential selection bias
Potential prognostic factors and how they were measured (A-E)	• Demographics factors: Age, Social deprivation: index of multiple deprivation, BMI, Smoker status, Work status: currently off work due to shoulder pain; not being in employment due to redundancy, unemployment or disability, compared with being in employment or education, Occupation: nature of employment; type of work or regular activity, Gender, Physical activity levels: most strenuous weekly exercise classified as 'none' compared to 'moderate' • Clinical history factors: Reason for symptom onset, Timing of onset of symptoms, History of previous shoulder pain, Paraesthesia in arm, Taking pain medication, Previously receiving treatment: physiotherapy helpful for previous shoulder problems, Side-related symptoms: presence of pain in the opposite upper quadrant; both shoulder affected or patient stated 'ambidextrous', Symptom duration, Co-morbidities: additional health problems • Patient reported measures: Pain (pain intensity), Psychological symptoms (Self-efficacy for pain, Patient expectation, Anxiety and depression in the previous 7 days), Function / Disability (SPADI, QuickDASH) • Clinical measures: Strength (shoulder force (abduction, ER)), ROM (reduced range of active shoulder abduction, increasing difference between the range of active and passive shoulder abduction, AROM (flexion), PROM (flexion, abduction, ER)), Scapular movement and control (change in shoulder pain/ range

	during manual facilitation of the scapula around the chest wall during arm elevation) • Structural pathology: Imaging (none), Orthopaedic tests (External lag sign)
Treatment (type; duration and frequency)	Type: entirely clinician directed; no details of how categorised
Outcome (how defined; timing; loss to follow up)	Disability via absolute Quick DASH score at 6 weeks post baseline and 6 months post baseline; Pain and disability via absolute SPADI score at 6 weeks post baseline and 6 months post baseline Loss to follow up: %; analysis of differences between groups = younger patients as being lost to follow up as well as those not partaking in leisure time physical activity
Analysis (how candidate variables selected, incl no cases per variable) (Multivariate regression analysis approach)	No attempt made to trim the number of candidate variables Number of cases per variable = 11.8 Most parsimonious model via forward selection and backward elimination R^2 value = 30%; 34%; 43%; 48%

Conroy and Hayes (1998)

Study design	Randomised control trial
Pathology; inclusion/exclusion	'Shoulder impingement syndrome' Subject selection criteria included pain about the superolateral shoulder region and one or more of the following findings: active range of motion deficits in humeral elevation, painful subacromial compression (Neer), Hawkins and Kennedy test. Exclusion: shoulder instability, primary scapule thoracic dysfunction, stage II and III adhesive capsulitis, third degree musculotendinous tears, advanced acromioclavicular joint disease, advanced calcific tendinitis or bursitis, severe degenerative bony or ligamentous changes, neurological involvement and unstable fracture of the humerus, scapula, or clavicle.
Patient care pathway and clinical setting	n/a
Sample size and nature of those not consenting	n=14; no details
Potential prognostic factors and how they were measured (A-E)	• Demographics factors: Age • Clinical history factors: Side-related symptoms: dominant arm involvement, Symptom duration • Patient reported measures: Pain (none) , Psychological symptoms (none), Function / Disability (none) • Clinical measures: Strength (none), ROM (none), Scapular movement and control (none) • Structural pathology: Imaging (none), Orthopaedic tests (none)
Treatment (type; duration and frequency)	Type: standardised interventions (muscle strengthening, joint mobilisation) Duration: 3 weeks; Frequency: three times a week
Outcome (how defined; timing; loss to follow up)	VAS pain score; upon completion of treatment No details
Analysis (how candidate variables selected, incl no cases per variable) (Multivariate regression analysis approach)	No attempt made to trim the number of candidate variables Two tailed, independent sample t tests or Mann-Whitney U tests

Deutscher et al. (2009)

Study design	Prospective cohort study
Pathology; inclusion/exclusion	Non-specific shoulder pain 'Shoulder pain'
Patient care pathway and clinical setting	Fifty-four community-based outpatient physical therapy clinics in Israel
Sample size and nature of those not consenting	n=5,000 No data
Potential prognostic factors and how they were measured (A-E)	• Demographics factors: Age, gender, smoker status, physical activity levels • Clinical history factors: Symptom duration: chronic symptoms, Taking antidepressant medication, Taking pain medication, Co-morbidities: number of co-morbidities • Patient reported measures: Pain (none), Psychological symptoms (none), Function / Disability (CAT) • Clinical measures: Strength (none), ROM (none), Scapular movement and control (none) • Structural pathology: Imaging (none), Orthopaedic tests (none)
Treatment (type; duration and frequency)	Type: entirely clinician directed; categorised using electronic database system
Response to intervention	n/a
Outcome (how defined; timing; loss to follow up)	Functional status via absolute CAT score at discharge; upon completion of treatment Loss to follow up: 60%; analysis of differences between groups = reported their missing cohort subjects as comprising a higher percentage of patients with a chronic condition at intake and a higher number of comorbidities
Analysis (how candidate variables selected, incl no cases per variable) (Multivariate regression analysis approach)	Reduced number of candidate variables by collapsing of variables within a category if distributions were too low in any one variable; the variable with the greatest clinical relevance was retained where collinearity amongst independent variables was identified. Most parsimonious model via forward selection and backward elimination R^2 value = 30%; 36%

Engebretsen et al. (2010)

Study design	Randomised control trial
Pathology; inclusion/exclusion	'Subacromial shoulder pain' Patients between 18 and 70 years old with subacromial shoulder pain for at least 3 months. The inclusion criteria were: dysfunction or pain on abduction; a normal passive glenohumeral range of motion; pain on two of three isometric tests (abduction at 0° or 30°, external or internal rotation); and a positive impingement sign. The exclusion criteria were: bilateral shoulder pain, previous surgery on the affected shoulder, instability, referred pain from neck, rheumatoid arthritis, clinical and radiological signs of glenohumeral- or acromioclavicular arthritis, serious somatic or psychiatric disorder or inability to understand Norwegian
Patient care pathway and clinical setting	Recruited from orthopaedic clinics and university hospital outpatient departments, Norway
Sample size and nature of those not consenting	n=104 Sample size calculations were presented these related to determining the sample required for differences to be identified in RCTs
Potential prognostic factors and how they were measured (A-E)	• Demographics factors: Age, gender, Work status, Occupation: frequency of working above shoulder height; frequency of carrying heavy loads at work, educational attainment • Clinical history factors: Recurrent shoulder pain: previous shoulder pain, Symptom duration, Associated neck pain, Side-related symptoms: dominant arm involvement, Previously receiving treatment: previous physiotherapy, Taking medication • Patient reported measures: Pain (pain intensity), Psychological symptoms (General health status, Emotional distress: Hopkins Symptom checklist, self-efficacy for pain), Function / Disability (SPADI) • Clinical measures: Strength (none), ROM (AROM – HBB, flexion on the affected side), Scapular movement and control (none) • Structural pathology: Imaging (none), Orthopaedic tests (none)
Treatment (type; duration and frequency)	Type: standardised interventions (treated with either supervised exercises or radial extracorporeal shockwave therapy or supervised exercises which emphasised relearning of normal movement patterns with an initial focus on unloading the rotator cuff via postural correction and manual techniques, patient education) Duration: maximum of 12 weeks; Frequency: once a week
Response to intervention	Worsening with physiotherapy = 3%
Outcome (how defined; timing; loss to follow up)	Pain and disability via absolute SPADI score; 'not working' status at 12 months post baseline Loss to follow up: 10%; analysis of differences between groups = missing subjects older and with a higher level of functional impairment at baseline compared to those for whom follow up at 1 year was possible
Analysis (how candidate variables selected, incl no cases per variable)	Reduced number of candidate variables via statistical methods – looking at univariate relationships between variables; the variable

(Multivariate regression analysis approach)	with the greatest clinical relevance was retained where collinearity amongst independent variables was identified. Number of cases per variable = 5.5 Age, gender and treatment group controlled for Most parsimonious model via variables with $p \geq 0.05$ removed Logistic regression R^2 value = 30%

Hung et al. (2010)

Study design	Prospective cohort study
Pathology; inclusion/exclusion	'Subacromial impingement' Subjects had to demonstrate at least 3 of the following: (1) a positive Neer impingement test, (2) a positive Hawkins impingement test, (3) a painful arc, (4 pain with isometric resisted abduction, (5) pain with palpation of the rotator cuff tendons, and (6) pain with active shoulder elevation. Subjects were excluded if they demonstrated signs of a complete rotator cuff tear or acute inflammation.
Patient care pathway and clinical setting	Recruited patients with SAIS from the orthopedics clinic in National Taiwan University Hospital and also through general announcements in local Internet media
Sample size and nature of those not consenting	n=33 Based on pilot study data, a sample size of 33 subjects was calculated to provide 80% power to detect differences of 6° of scapular kinematic variables with 6° standard deviation between the improvement and non-improvement groups 58 patients with SAIS recruited; after screening of the patients with the tests (criteria), 33 subjects met the criteria for the study. No details on potential bias
Potential prognostic factors and how they were measured (A-E)	• Demographics factors: Age • Clinical history factors: None • Patient reported measures: Pain (none), Psychological symptoms (none), Function / Disability (FLEX-SF) • Clinical measures: Strength (scapular protractor force, ER, IR and abduction), ROM (PROM IR, ER; posterior shoulder tightness (PST)), Scapular movement and control (Scapular kinematics (UR, IR and scapular tilt) at 30°, 60°, 90°, 120° elevation during both the ascending and descending phases in loaded and unloaded conditions) • Structural pathology: Imaging (none), Orthopaedic tests (none)
Treatment (type; duration and frequency)	Type: stretching, strengthening and ROM exercises along with manual therapy Duration: 6 weeks; Frequency: twice a week
Response to intervention	Successful outcome = 70%
Outcome (how defined; timing; loss to follow up)	'Improvement' via GROC score => 4 at discharge; upon completion of treatment Loss to follow up: 3% but failed to compare the baseline characteristics of those who were and those who were not followed up
Analysis (how candidate variables selected, incl no cases per variable) (Multivariate regression analysis approach)	Reduced number of candidate variables via statistical methods – looking at univariate relationships between variables Number of cases per variable = 0.5 Most parsimonious model via forward selection and backward elimination Logistic regression R^2 values not presented

Kennedy et al. (2006)

Study design	Prospective cohort study
Pathology; inclusion/exclusion	Non-specific shoulder pain Soft tissue shoulder disorders: Shoulder complaints were defined as any condition of pain or discomfort, including instances where there had been surgical treatment of the soft tissue shoulder disorder (eg, rotator cuff repair). 8% post-surgery. Patients were excluded from the study if they: (1) had fractures or dislocations associated with soft tissue pain, (2) received physical therapy for only one visit (eg, referred for equipment or single education session), or (3) were unable to read and write English and thus could not complete the questionnaire package independently
Patient care pathway and clinical setting	Patients beginning treatment at a large number of different physiotherapy practices; Canada
Sample size and nature of those not consenting	n=361 Yes - compared demographic variables describing participants and those considered nonparticipants (exclusion by eligibility criteria or the patient refused to participate) and did not find any statistically significant differences between the 2 groups. There were no sex and age differences between the participants group and nonparticipants group (P >= 0.4). Although nonparticipants had their symptoms for a longer time before starting therapy than participants (381 versus 229 days), the difference was marginally (non)significant (unpaired t test, P=.07)
Potential prognostic factors and how they were measured (A-E)	• Demographics factors: Age, gender, Work status, workers compensation claim • Clinical history factors: Symptom duration: shorter duration in symptoms, Post-surgical case, Recurrent shoulder pain, Reason for symptom onset, Co-morbidities: number of co-morbidities and number that limit activity, Taking pain medication • Patient reported measures: Pain (pain intensity), Psychological symptoms (Patient expectation, Mental Component Score (MCS) from the 36-Item Short-Form Health Survey (SF-36)), Function / Disability (DASH, Physical Component Score (PCS) from the 36-Item Short-Form Health Survey (SF-36), Global rating of disability score) • Clinical measures: Strength (muscle wasting and muscle weakness), ROM (active and passive ROM), Scapular movement and control (none) • Structural pathology: Imaging (none), Orthopaedic tests (none)
Treatment (type; duration and frequency)	Type: no details Mean 14.8 sessions (SD=8.7), duration of Rx = mean of 64.9 days (SD=25.6).
Response to intervention	Worsening with physiotherapy = 7%
Outcome (how defined; timing; loss to follow up)	Disability via absolute DASH score at discharge / 12 weeks; change in disability via change from DASH score at baseline to discharge / 12 weeks Loss to follow up: failed to compare the baseline characteristics of those who were and those who were not followed up

Analysis (how candidate variables selected, incl no cases per variable) (Multivariate regression analysis approach)	Reduced number of candidate variables via statistical methods – looking at univariate relationships between variables; also collapsing of variables within a category if distributions were too low in any one variable; the variable with the greatest clinical relevance was retained where collinearity amongst independent variables was identified. Number of cases per variable = 10 Most parsimonious model via variables with $p \geq 0.05$ removed R^2 value = 23%

Kromer et al. (2014)

Study design	Randomised control trial
Pathology; inclusion/exclusion	'Subacromial pain syndrome' Inclusion criteria set for this trial were: (1) age between 18 and 75 years; (2) symptoms for at least 4 weeks; (3) main complaints in the glenohumeral joint region or the proximal segments of the arm; (4) presence of one of the following signs indicating SPS: Neer impingement sign, Hawkins- Kennedy impingement test, or painful arc with active abduction or flexion; and (5) pain during one of the following resistance tests: external rotation, internal rotation, abduction, or flexion. Exclusion criteria were: (1) average 24-hour pain of 8/10 or more on a visual numeric rating scale (VNRS); (2) primary scapulothoracic dysfunction due to paresis; (3) diagnosed instability or previous history of dislocation; (4) frozen shoulder; (5) more than one-third restriction of elevation compared with the unaffected side; (6) substantial shoulder weakness or loss of active shoulder function; (7) shoulder surgery on the involved side in the previous 12 months; (8) reproduction of symptoms with active or passive cervical movements; (9) neurological involvement with sensory or muscular deficits; (10) inflammatory joint disease (eg, rheumatoid arthritis); (11) diabetes mellitus; (12) intake of psychotherapeutic drugs; (13) compensation claims; and (14) inability to understand written or spoken German.
Patient care pathway and clinical setting	Germany, outpatient. Presentation to a physical therapist following referral by general practitioner or orthopedic surgeon (duration of symptoms >=4 wk) Outpatient physiotherapy clinics, GPs and orthopaedic surgeons
Sample size and nature of those not consenting	n=90; sample size calculations were presented these related to determining the sample required for differences to be identified in RCTs
Potential prognostic factors and how they were measured (A-E)	• Demographics factors: Age and gender were controlled for in the regression model • Clinical history factors: Symptom duration • Patient reported measures: Pain (pain intensity), Psychological symptoms (Fear Avoidance Beliefs Questionnaire (physical activity subscale), Pain catastrophising scale), Function / Disability (Function / disability as measured by the function element of the SPADI) • Clinical measures: Strength (none), ROM (none), Scapular movement and control (none) • Structural pathology: Imaging (none), Orthopaedic tests (none)
Treatment (type; duration and frequency)	Type: standardised interventions (shoulder girdle and thoracic spine stretching and strengthening exercises, manual mobilisation techniques for the shoulder complex and cervical spine, patient education) Duration: 5 weeks of clinician directed rehabilitation followed by 7 weeks of home based exercises; Frequency: twice a week

Outcome (how defined; timing; loss to follow up)	Change in functional status via change in functional subscale of SPADI (SPADI-F) between baseline and 3 months post baseline Loss to follow up: 2% but failed to compare the baseline characteristics of those who were and those who were not followed up
Analysis (how candidate variables selected, incl no cases per variable) (Multivariate regression analysis approach)	Number of cases per variable = 18 Reduced number of candidate variables via variable with the greatest clinical relevance was retained where collinearity amongst independent variables was identified. Hierarchical regression: first step (variables (age and gender) controlling for), second step explored clinical variables and third step explored psychological variables Most parsimonious model via variables with $p \geq 0.05$ removed R^2 value = 48%

Mintken et al. (2010)

Study design	Prospective cohort study
Pathology; inclusion/exclusion	Non-specific shoulder pain Inclusion criteria: between ages of 18 and 65 years, with a primary report of shoulder pain and a baseline Shoulder Pain and Disability Index (SPADI) score of 20% or greater. Exclusion criteria: any medical "red flags" suggestive of a non-musculoskeletal aetiology of symptoms, acute fractures in the shoulder region, acute severe trauma in the cervical or thoracic region in the previous 6 weeks, a diagnosis of cervical spinal stenosis or bilateral upper-extremity symptoms, osteoporosis, prior surgery to the cervical or thoracic region, evidence of central nervous system involvement, insufficient English language skills to complete the questionnaires, or signs consistent with nerve root compression (defined as impairment in at least 2 of the following: myotomal strength, sensation, or reflexes). "Red flags" were ruled out by a combination of a medical screening questionnaire, a neurological examination, and a patient history.
Patient care pathway and clinical setting	Outpatient physiotherapy clinics
Sample size and nature of those not consenting	n=80 No details on n=8 who declined to participate
Potential prognostic factors and how they were measured (A-E)	• Demographics factors: Age, gender, work status • Clinical history factors: Symptom duration, Taking pain medication, Reason for symptom onset, Recurrent shoulder pain: number of previous episodes, Previously receiving treatment: treatment for previous episodes, past medical history • Patient reported measures: Pain (pain intensity), Psychological symptoms (Patient expectation, Fear Avoidance Beliefs Questionnaire, Tampa Scale for Kinesiophobia), Function / Disability (SPADI) • Clinical measures: Strength (Serratus Anterior, Middle trapezius, Lower trapezius, Rhomboid, Deltoid, External and internal shoulder rotator muscle strength), ROM (pain-free shoulder flexion, passive shoulder IR at 90° abduction, Passive shoulder abduction, Passive shoulder ER at 90° abduction, Battery of 3 functional tests), Scapular movement and control (Lateral slide test and scapula index, Qualitative assessment of scapular function) • Structural pathology: Imaging (none), Orthopaedic tests (13 orthopaedic tests including the Hawkins-Kennedy impingement test, the empty can and full can test and the drop sign Neer's sign)
Treatment (type; duration and frequency)	Type: thrust manipulations of the cervicothoracic spine Duration: days
Response to intervention	Successful outcome = 61%; Worsening with physiotherapy = 4%
Outcome (how defined; timing; loss to follow up)	'Improvement' via GROC score => 4 at discharge; upon completion of treatment

	At the beginning of the second session, the participants completed the GROC and the other outcome measures. If their score on the GROC did not exceed the +4 cutoff at the second session, they received the same intervention program again and were scheduled for a follow-up within 2 to 4 days. Participants again completed the GROC along with the other outcome measures. If they scored +4 or better on the GROC, they were categorized as having a successful outcome; if they scored below +4, they were categorized as not having a successful outcome. At this point, their participation in the study was complete, and the therapist could administer further treatment as needed. F/up was at 2nd or 3rd apt – over a period of a few days Loss to f/u = 1
Analysis (how candidate variables selected, incl no cases per variable) (Multivariate regression analysis approach)	Reduced number of candidate variables via statistical methods – looking at univariate relationships between variables Variables with a significance level of P<.10 were retained as potential prognostic variables. For continuous variables with a significant univariate relationship, sensitivity and specificity values were calculated for all possible cutoff points and then plotted as a receiver operator characteristic (ROC) curve. Number of cases per variable = 6 Logistic regression

Morrison et al. (1997)

Study design	Retrospective cohort study
Pathology; inclusion/exclusion	'Subacromial impingement' Diagnosis made on the basis of a positive impingement sign and the absence of other abnormalities of the shoulder, such as full thickness tears of the rotator cuff, osteoarthrosis of the acromioclavicular joint, instability of the glenohumeral joint, or adhesive capsulitis. Diagnosis made on the basis of a history, a clinical examination, and a positive Neer impingement sign. Patients who had concomitant adhesive capsulitis, cervical radiculopathy, or suprascapular nerve palsy were excluded, as were patients who had major weakness on testing of the rotator cuff or a full-thickness tear of the rotator cuff on magnetic resonance imaging.
Patient care pathway and clinical setting	Investigation performed at the Southern California Centre for Sports Medicine, Long Beach
Sample size and nature of those not consenting	n=616 No data
Potential prognostic factors and how they were measured (A-E)	• Demographics factors: Age, gender • Clinical history factors: Symptom duration, Side-related symptoms: dominant arm involvement • Patient reported measures: Pain (none), Psychological symptoms (none), Function / Disability (none) • Clinical measures: Strength (none), ROM (none), Scapular movement and control (none) • Structural pathology: Imaging (Acromion morphology type 1 on X-ray), Orthopaedic tests (none)
Treatment (type; duration and frequency)	Type: physiotherapy programme consisting of stretching and then strengthening
Response to intervention	Successful outcome = 67%
Outcome (how defined; timing; loss to follow up)	Pain, function, active ROM, strength and overall satisfaction via absolute UCLA score at average of 27 months Loss to follow up: 8 % but failed to compare the baseline characteristics of those who were and those who were not followed up
Analysis (how candidate variables selected, incl no cases per variable) (Multivariate regression analysis approach)	No attempt made to trim the number of candidate variables Chi-square Number of cases per variable = 123

Ogon et al. (2009)

Study design	Prospective cohort study
Pathology; inclusion/exclusion	'Calcific tendonitis' Inclusion criteria were the presence of radiographically and sonographically proven calcific deposits in a rotator cuff tendon and the presence of clinically symptomatic calcific tendinitis of the shoulder requiring continuation of treatment at the time of presentation at the institution. Exclusion criteria were previous surgical interventions, needling, application of ultrasound therapy, or extracorporeal shock wave therapy (ESWT), as well as the presence of rheumatoid arthritis or concomitant diseases of the affected shoulder
Patient care pathway and clinical setting	All patients had previously received nonoperative treatment from general practitioners, rheumatologists, or orthopedic surgeons, which included physical therapy, manual therapy, electrotherapy, iontophoresis, systemic use of analgesics and nonsteroidal antiinflammatory drugs (NSAIDs), and up to 3 subacromial injections of corticosteroids. The patients were then referred to the Orthopaedic Outpatient Clinic because of the persistence of clinically symptomatic calcific tendinitis of the shoulder.
Sample size and nature of those not consenting	N=420 No data
Potential prognostic factors and how they were measured (A-E)	• Demographics factors: Age, gender, occupation • Clinical history factors: Periods of professional disability, Reason for symptom onset, Symptom duration, Side-related symptoms: dominant arm involvement, Past medical history • Patient reported measures: Pain (none), Psychological symptoms (none), Function / Disability (none) • Clinical measures: Strength (none), ROM (none), Scapular movement and control (none) • Structural pathology: Imaging (Calcific deposits: X-ray and sonographic), Orthopaedic tests (none)
Treatment (type; duration and frequency)	Type: physiotherapy treatment algorithm including heat or cold and manual therapy Duration: minimum of 3 months
Response to intervention	Successful outcome = 73%
Outcome (how defined; timing; loss to follow up)	'Success of non-operative therapy' via no progression to advanced therapeutic measures after a minimum of 6 months non-operative treatment (including minimum of 3 months treatment at the study location) Loss to follow up: failed to compare the baseline characteristics of those who were and those who were not followed up
Analysis (how candidate variables selected, incl no cases per variable) (Multivariate regression analysis approach)	No attempt made to trim the number of candidate variables Number of cases per variable = 25 Most parsimonious model via variables with $p \geq 0.05$ removed Prognostic factors were determined at $P < 0.05$ by chi-square test Logistic regression R^2 values not presented

Sindhu et al. (2012)

Study design	Retrospective cohort study
Pathology; inclusion/exclusion	Non-specific shoulder pain Clinical staff entered necessary medical information at intake, such as diagnosis codes based on the International Classification of Diseases, Ninth Revision (ICD-9).
Patient care pathway and clinical setting	People with musculoskeletal conditions of the shoulder who attended outpatient rehabilitation clinics throughout the United States Data were collected from people with musculoskeletal conditions of the shoulder receiving rehabilitation.
Sample size and nature of those not consenting	N=3362 N/a
Potential prognostic factors and how they were measured (A-E)	• Demographics factors: None • Clinical history factors: None • Patient reported measures: Pain (none), Psychological symptoms (Fear Avoidance Beliefs Questionnaire), Function / Disability (none) • Clinical measures: Strength (none), ROM (none), Scapular movement and control (none) • Structural pathology: Imaging (none), Orthopaedic tests (none)
Treatment (type; duration and frequency)	Type: no details
Outcome (how defined; timing; loss to follow up)	Functional change via change in CAT score between baseline and discharge; upon completion of treatment Loss to follow up: 43%; analysis of differences between groups = identified differences between those available for and lost to follow up in terms of age, pain levels and function but provided no details on the direction of difference
Analysis (how candidate variables selected, incl no cases per variable) (Multivariate regression analysis approach)	General linear model (GLM) used to describe how change in function is affected by fear avoidance 3362

Tyler et al. (2010)

Study design	Prospective cohort study
Pathology; inclusion/exclusion	'Internal impingement' Diagnostic criteria for internal impingement was used. Specific inclusion criteria based on physical examination were positive relocation test, positive posterior impingement sign, and posterior glenohumeral joint line tenderness. Specific inclusion criteria based on MRI findings were the presence of a posterosuperior glenoid labral lesion. Exclusion criteria were anterior instability, full-thickness rotator cuff tear, and subacromial impingement as determined by physical examination and MRI. Additionally, all patients reported subjective clicking in their shoulder on active movement.
Patient care pathway and clinical setting	No details
Sample size and nature of those not consenting	n=22
Potential prognostic factors and how they were measured (A-E)	• Demographics factors: None • Clinical history factors: None • Patient reported measures: Pain (none), Psychological symptoms (none), Function / Disability (none) • Clinical measures: Strength (none), ROM (Improvement in PST at discharge, Improvement in passive ER ROM, Improvement in GIRD), Scapular movement and control (none) • Structural pathology: Imaging (none), Orthopaedic tests (none)
Treatment (type; duration and frequency)	Type: glenohumeral joint glides, sleeper stretches and cross-chest adduction Duration 7 ± 2 weeks; Frequency: three times a week
Outcome (how defined; timing; loss to follow up)	Symptom free via Simple Shoulder Test at discharge; upon completion of treatment
Analysis (how candidate variables selected, incl no cases per variable) (Multivariate regression analysis approach)	No attempt made to trim the number of candidate variables Number of cases per variable = 7 Mixed model analysis of variance, with Treatment (pretreatment vs posttreatment) as the within-subjects factor and Group (patients with complete resolution of symptoms vs patients with residual symptoms) as the between-subjects factor.

Virta et al. (2009)

Study design	Prospective cohort study
Pathology; inclusion/exclusion	'Subacromial impingement' Diagnosis was confirmed with subacromial anaesthesia, with a few exceptions by the same doctor, and most of the patients had passed an MRI examination. Patients were included if diagnosis was confirmed
Patient care pathway and clinical setting	Recruited patients who were on a waiting list for orthopaedic surgery but had been referred for physiotherapy prior to surgery
Sample size and nature of those not consenting	n=97
Potential prognostic factors and how they were measured (A-E)	• Demographics factors: Age, gender • Clinical history factors: Symptom duration • Patient reported measures: Pain (none), Psychological symptoms (none), Function / Disability (none) • Clinical measures: Strength (none), ROM (none), Scapular movement and control (none) • Structural pathology: Imaging (none), Orthopaedic tests (none)
Treatment (type; duration and frequency)	Type: ROM and strengthening exercises Duration: 8 weeks; Frequency: twice a week
Response to intervention	Successful outcome = 87%
Outcome (how defined; timing; loss to follow up)	UCLA; upon completion of treatment Loss to follow up: 26% but failed to compare the baseline characteristics of those who were and those who were not followed up
Analysis (how candidate variables selected, incl no cases per variable) (Multivariate regression analysis approach)	No attempt made to trim the number of candidate variables No details of analysis Number of cases per variable = 18 R^2 values not presented

10.2 PROBAST process

The following is adapted from PROBAST: Prediction Risk of Bias Tool 2017. York, United Kingdom: Kleijnen Systematic Reviews Ltd. 2017. Available at: http://www.systematic-reviews.com/ probast. Accessed October 6, 2017.

PROBAST includes five steps.

Step	Task	When to complete
1	Specify your systematic review question	Once per systematic review
2	Classify the type of prediction model evaluation	Once for each model of interest in each publication being assessed, for each relevant outcome
3	Assess risk of bias and applicability	Once for each evaluation (development and/or validation) of each distinct model
4	Overall judgement	Once for each evaluation (development and/or validation) of each distinct model
5	Usability of the model	Once for each distinct model

Step 1: Specify your systematic review question

State your systematic review question to facilitate the assessment of the applicability of the evaluated models to your question. The following table should be completed once per systematic review.

Specify your systematic review question
Participants (e.g. setting, main inclusion criteria, prior treatments):
Outcome(s) to be predicted:
Intended use of the model(s): **When will the model(s) be used, e.g. at presentation with signs/ symptoms, staging severity of disease, pre-operatively?**

Step 2: Classify the type of prediction model evaluation

Different signalling questions apply for different types of prediction model evaluation. Use the following table to classify the evaluation as model development, model validation or both. If the evaluation does not fit one of these classifications then PROBAST should not be used.

Type of model evaluation	Tick as appropriate	PROBAST classification

Step 3: Assess risk of bias and applicability

PROBAST is structured as five key domains. Each domain is judged for risk of bias (low, high or unclear) and includes signalling questions to help make judgements. Signalling questions are rated as yes (Y), probably yes (PY), probably no (PN), no (N) or no information (NI). All signalling questions are

phrased so that “yes” indicates absence of bias. Any signalling question rated as “no” or “probably no” flags the potential for bias; you will need to use your judgement to determine whether the domain should be rated as “high”, “low” or “unclear” risk of bias.

DOMAIN 1: **Participant selection**

A. Risk of Bias

Describe the sources of data and criteria for participant selection:	
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	
2. Were all inclusions and exclusions of participants appropriate?	
3. Were participants enrolled at a similar state of health, or were predictors considered to account for any dissimilarities?	
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)

B. Applicability

Describe included participants, setting and dates:	
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)

DOMAIN 2: **Predictors**

A. Risk of Bias

List and describe predictors assessed, e.g. definition, timing and knowledge of other predictors:	
1. Were predictors defined and assessed in a similar way for all participants?	
3. Were predictor assessments made without knowledge of outcome data?	
4. Are all predictors available at the time the model is intended to be used?	
5. Were all relevant predictors analysed?	
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)

Note: question 2 pertained to model validation studies only

B. Applicability

Concern that the definition, assessment or timing of assessment of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)

DOMAIN 3: **Outcome**

A. Risk of Bias

Describe the outcome and how it was defined and determined:	
1. Was a pre-specified outcome definition used?	
2. Were predictors excluded from the outcome definition?	
3. Was the outcome defined and determined in a similar way for all participants?	
5. Was the outcome determined without knowledge of predictor information?	
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)

Note: question 4 pertained to model validation studies only

B. Applicability

At what time point was the outcome determined: If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:	
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)

DOMAIN 4: **Sample size and participant flow**

A. Risk of Bias

Describe numbers of participants, outcome events and events per predictor: Describe the time interval between predictor assessment and outcome determination: Describe any participants who were excluded from the model: Describe missing data on predictors and outcomes as well as methods used for missing data:

1. Were there a reasonable number of outcome events?	
2. Was the time interval between predictor assessment and outcome determination appropriate?	
3. Were all enrolled participants included in the analysis?	
4. Were participants with missing data handled appropriately?	
Risk of bias introduced by sample size or participant flow	RISK: (low/ high/ unclear)

DOMAIN 5: **Analysis**

A. Risk of Bias

Describe how the model was developed (predictor selection, fitting and optimism, risk groups, model performance): Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): Describe the performance measures of the model, e.g. calibration, discrimination, (re)classification, net benefit:	
1. Were non-binary predictors handled appropriately?	
2. Was selection of predictors based on univariable analysis avoided?	
3. Was model overfitting (optimism in model performance) accounted for, e.g. using bootstrapping or shrinkage techniques?	
4. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	
6. Were relevant model performance measures evaluated, e.g. calibration, discrimination, (re)classification and net benefit?	
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)

Note: question 7 pertained to model validation studies only

Step 4: Overall judgement

Use the following tables to reach overall judgements about risk of bias and applicability of the prediction model evaluation (development and/ or validation) across all assessed domains. Complete for each evaluation of a distinct model.

Reaching an overall judgement about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as low risk of bias for all domains, consider downgrading to high risk of bias. Such model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias.
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Reaching an overall judgement about applicability of the prediction model evaluation	
Low concerns regarding applicability	If low concerns regarding applicability for all domains, the prediction model evaluation is judged to have low concerns regarding applicability.
High concerns regarding applicability	If high concerns regarding applicability for at least one domain, the prediction model evaluation is judged to have high concerns regarding applicability.
Unclear concerns regarding applicability	If unclear concerns (but no "high concern") regarding applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns regarding applicability overall.

Overall judgement about risk of bias and applicability of the prediction model evaluation	
Overall judgement of risk of bias	RISK: (low/ high/ unclear)
Overall judgement of applicability	CONCERN: (low/ high/ unclear)

Step 5: Usability of the model

The following question assesses whether the model was presented in enough detail to be usable in the targeted individuals and context. Note that this is different from the applicability assessment above, which refers to the extent to which the prediction model evaluation matches your review question.

Complete for each evaluation of a distinct model or simplified score.

Assess the usability of the model	
Is the model presented with sufficient detail to be used in the intended context and target population?	RATING: (yes/ no)

10.3 New 10.3 PROBAST Outcomes

Step 1: Specify your systematic review question

Specify your systematic review question What are the prognostic indicators of successful rehabilitation outcome in patients with Subacromial Impingement Syndrome / Rotator Cuff Tendinopathy?
Participants (e.g. setting, main inclusion criteria, prior treatments): Patients in the Cardiff & Vale region who have been referred for out-patient physiotherapy in the NHS, secondary care setting. Patients with a clinical diagnosis of Subacromial Impingement Syndrome / Rotator Cuff Tendinopathy. No recent shoulder surgery; otherwise no restriction placed upon previous treatment. Treatment defined as non-invasive, multimodal physiotherapy.
Outcome(s) to be predicted: Change in shoulder function from baseline.
Intended use of the model(s): Prognosis **When will the model(s) be used, e.g. at presentation with signs/ symptoms, staging severity of disease, pre-operatively?** Upon referral to NHS out-patient physiotherapy in a secondary care setting.

Step 2: Classify the type of prediction model evaluation

Type of model evaluation	Tick as appropriate	PROBAST classification
Prediction model development without testing its predictive performance in other individuals, i.e. no external validation. Model development should ideally include internal validation, such as bootstrapping or cross-validation.	All studies included in the review were of this type	Development only

Bartolozzi et al. (1994)

Step 3: Assess risk of bias and applicability

DOMAIN 1: **Participant selection**

A. Risk of Bias

Describe the sources of data and criteria for participant selection: • Patients attending orthopaedic department with impingement syndrome treated non-operatively • All pts had X-rays, 101 had MRI; for other 35 patients diagnosis was based entirely on history and clinical examination – comprising painful arc, +ve impingement sign, +ve impingement test, specific rotator cuff weakness. None of the 35 had a +ve drop arm test or clinically evident RC weakness.	
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Yes

2. Were all inclusions and exclusions of participants appropriate? Diagnosis not typically made based on X-ray and MRI imaging in isolation	No
3. Were participants enrolled at a similar state of health, or were predictors considered to account for any dissimilarities?	Unclear
Risk of bias introduced by selection of participants	RISK: High

B. Applicability

Describe included participants and setting: n=136 Patients attending orthopaedic department with impingement syndrome treated non-operatively	
Concern that the included participants and setting do not match the review question	CONCERN: Low

DOMAIN 2: **Predictors**

A. Risk of Bias

List and describe predictors assessed, e.g. definition, timing and knowledge of other predictors: • Demographics factors: Age, gender, occupation • Clinical history factors: Side-related symptoms: dominant arm involvement, Reason for symptom onset, Symptom duration • Patient reported measures: Pain (none), Psychological symptoms (none), Function / Disability (UCLA) • Clinical measures: Strength (shoulder girdle weakness), ROM (active abduction and flexion ROM), Scapular movement and control (none) • Structural pathology: Imaging (MRI: Rotator cuff pathology), Orthopaedic tests (none) All recorded at initiation of the study	
1. Were predictors defined and assessed in a similar way for all participants?	Yes
3. Were predictor assessments made without knowledge of outcome data?	Yes
4. Are all predictors available at the time the model is intended to be used?	Yes
5. Were all relevant predictors analysed?	Yes
Risk of bias introduced by predictors or their assessment	RISK: Low

B. Applicability

Concern that the definition, assessment or timing of assessment of predictors in the model do not match the review question	CONCERN: Low

DOMAIN 3: **Outcome**

A. Risk of Bias

Describe the outcome and how it was defined and determined: Combination of UCLA score at discharge and perceived improvement	
1. Was a pre-specified outcome definition used?	Yes
2. Were predictors excluded from the outcome definition? UCLA both predictor and outcome	No
3. Was the outcome defined and determined in a similar way for all participants?	Yes
5. Was the outcome determined without knowledge of predictor information?	Yes
Risk of bias introduced by the outcome or its determination	RISK: High

B. Applicability

At what time point was the outcome determined: At average of 20 months If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome: n/a	
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: Low

DOMAIN 4: **Sample size and participant flow**

A. Risk of Bias

Describe numbers of participants and events per predictor: • n=136; 10 Describe the time interval between predictor assessment and outcome determination: • Average of 20 months Describe any participants who were excluded from the model: • unknown Describe missing data on predictors and outcomes as well as methods used for missing data: • n/a	
1. Were there a reasonable number of outcome events?	Yes
2. Was the time interval between predictor assessment and outcome determination appropriate? Overly long time from initiation of study	No
3. Were all enrolled participants included in the analysis?	NI
4. Were participants with missing data handled appropriately?	NI
Risk of bias introduced by sample size or participant flow	RISK: High

DOMAIN 5: **Analysis**

A. Risk of Bias

Describe how the model was developed (predictor selection, fitting and optimism, risk groups, model performance): • No attempt made to trim the number of candidate variables • No details on model construction Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): • n/a Describe the performance measures of the model, e.g. calibration, discrimination, (re)classification, net benefit: • n/a	
1. Were non-binary predictors handled appropriately?	NI
2. Was selection of predictors based on univariable analysis avoided?	NI
3. Was model overfitting (optimism in model performance) accounted for, e.g. using bootstrapping or shrinkage techniques?	No
4. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	No
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	No
6. Were relevant model performance measures evaluated, e.g. calibration, discrimination, (re)classification and net benefit?	No
Risk of bias introduced by the analysis	RISK: High

Step 4: Overall judgement

Overall judgement about risk of bias and applicability of the prediction model evaluation	
Overall judgement of risk of bias	RISK: High
Overall judgement of applicability	CONCERN: Low

Step 5: Usability of the model

Assess the usability of the model	
Is the model presented with sufficient detail to be used in the intended context and target population?	RATING: No

Chester et al (2016)

Step 3: Assess risk of bias and applicability

DOMAIN 1: **Participant selection**

A. Risk of Bias

Describe the sources of data and criteria for participant selection: • Patients referred to physiotherapy for the management of musculoskeletal shoulder pain at 11 NHS trusts and social enterprises in the East of England were recruited. Participating physiotherapy departments were located within primary and secondary care • Patients were eligible if they were aged 18 years or older and described shoulder or arm pain aggravated by shoulder movements. Patients with significant reproduction of shoulder pain on spinal movement, or greater reproduction on spinal movement compared to shoulder movement, were excluded from the study. Patients with the following aetiology for shoulder pain were excluded: radiculopathy, postsurgery, postfracture, posttraumatic dislocation or systemic source.	
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Yes
2. Were all inclusions and exclusions of participants appropriate?	Yes
3. Were participants enrolled at a similar state of health, or were predictors considered to account for any dissimilarities?	Unclear
Risk of bias introduced by selection of participants	RISK: Low

B. Applicability

Describe included participants and setting: n=1030 Patients referred to physiotherapy for the management of musculoskeletal shoulder pain at 11 NHS trusts and social enterprises in the East of England were recruited. Participating physiotherapy departments were located within primary and secondary care	
Concern that the included participants and setting do not match the review question	CONCERN: High

DOMAIN 2: **Predictors**

A. Risk of Bias

List and describe predictors assessed, e.g. definition, timing and knowledge of other predictors: • Demographics factors: Age, Social deprivation: index of multiple deprivation, BMI, Smoker status, Work status: currently off work due to shoulder pain; not being in employment due to redundancy, unemployment or disability, compared with being in employment or education, Occupation: nature of employment; type of work or regular activity, Gender, Physical activity levels: most strenuous weekly exercise classified as 'none' compared to 'moderate' • Clinical history factors: Reason for symptom onset, Timing of onset of symptoms, History of previous shoulder pain, Paraesthesia in arm, Taking pain medication, Previously receiving treatment: physiotherapy helpful for previous shoulder problems, Side-related symptoms:

presence of pain in the opposite upper quadrant; both shoulder affected or patient stated 'ambidextrous', Symptom duration, Co-morbidities: additional health problems • Patient reported measures: Pain (pain intensity), Psychological symptoms (Self-efficacy for pain, Patient expectation, Anxiety and depression in the previous 7 days), Function / Disability (SPADI, QuickDASH) • Clinical measures: Strength (shoulder force (abduction, ER)), ROM (reduced range of active shoulder abduction, increasing difference between the range of active and passive shoulder abduction, AROM (flexion), PROM (flexion, abduction, ER)), Scapular movement and control (change in shoulder pain/ range during manual facilitation of the scapula around the chest wall during arm elevation) • Structural pathology: Imaging (none), Orthopaedic tests (External lag sign) All recorded at initiation of the study	
1. Were predictors defined and assessed in a similar way for all participants?	Yes
3. Were predictor assessments made without knowledge of outcome data?	Yes
4. Are all predictors available at the time the model is intended to be used?	Yes
5. Were all relevant predictors analysed?	Yes
Risk of bias introduced by predictors or their assessment	RISK: Low

B. Applicability

Concern that the definition, assessment or timing of assessment of predictors in the model do not match the review question	CONCERN: Low

DOMAIN 3: **Outcome**

A. Risk of Bias

Describe the outcome and how it was defined and determined: Disability via absolute Quick DASH score at 6 weeks post baseline and 6 months post baseline; Pain and disability via absolute SPADI score at 6 weeks post baseline and 6 months post baseline	
1. Was a pre-specified outcome definition used?	Yes
2. Were predictors excluded from the outcome definition? SPADI and QuickDASH were both predictors and outcome determinants	No
3. Was the outcome defined and determined in a similar way for all participants?	Yes
5. Was the outcome determined without knowledge of predictor information?	Yes
Risk of bias introduced by the outcome or its determination	RISK: High

B. Applicability

At what time point was the outcome determined: 6 weeks post baseline and 6 months post baseline If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome: n/a	
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: Low

DOMAIN 4: **Sample size and participant flow**

A. Risk of Bias

Describe numbers of participants and events per predictor: • n=1030; 11.8 Describe the time interval between predictor assessment and outcome determination: • 6 weeks post baseline and 6 months post baseline Describe any participants who were excluded from the model: • Younger patients as being lost to follow up as well as those not partaking in leisure time physical activity Describe missing data on predictors and outcomes as well as methods used for missing data:	
1. Were there a reasonable number of outcome events?	Yes
2. Was the time interval between predictor assessment and outcome determination appropriate?	Yes
3. Were all enrolled participants included in the analysis?	Yes
4. Were participants with missing data handled appropriately?	n/a
Risk of bias introduced by sample size or participant flow	RISK: Low

DOMAIN 5: **Analysis**

A. Risk of Bias

Describe how the model was developed (predictor selection, fitting and optimism, risk groups, model performance): • No attempt made to trim the number of candidate variables • Most parsimonious model via forward selection and backward elimination Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): • n/a Describe the performance measures of the model, e.g. calibration, discrimination, (re)classification, net benefit: • n/a

1. Were non-binary predictors handled appropriately?	Yes
2. Was selection of predictors based on univariable analysis avoided?	Yes
3. Was model overfitting (optimism in model performance) accounted for, e.g. using bootstrapping or shrinkage techniques?	No
4. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	No
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	Yes
6. Were relevant model performance measures evaluated, e.g. calibration, discrimination, (re)classification and net benefit?	No
Risk of bias introduced by the analysis	RISK: High

Step 4: Overall judgement

Overall judgement about risk of bias and applicability of the prediction model evaluation	
Overall judgement of risk of bias	RISK: High
Overall judgement of applicability	CONCERN: High

Step 5: Usability of the model

Assess the usability of the model	
Is the model presented with sufficient detail to be used in the intended context and target population?	RATING: No

Conroy and Hayes (1998)

Step 3: Assess risk of bias and applicability

DOMAIN 1: **Participant selection**

A. Risk of Bias

Describe the sources of data and criteria for participant selection:

- No data on care setting
- Subject selection criteria included pain about the superolateral shoulder region and one or more of the following findings: active range of motion deficits in humeral elevation, painful subacromial compression (Neer), Hawkins and Kennedy test.
  Exclusion: shoulder instability, primary scapule thoracic dysfunction, stage II and III adhesive capsulitis, third degree musculotendinous tears, advanced acromioclavicular joint disease, advanced calcific tendinitis or bursitis, severe degenerative bony or ligamentous changes, neurological involvement and unstable fracture of the humerus, scapula, or clavicle.

1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Yes
2. Were all inclusions and exclusions of participants appropriate?	Yes
3. Were participants enrolled at a similar state of health, or were predictors considered to account for any dissimilarities?	Unclear
Risk of bias introduced by selection of participants	RISK: Low

B. Applicability

Describe included participants and setting: n=14 No data on care setting	
Concern that the included participants and setting do not match the review question	CONCERN: High

DOMAIN 2: **Predictors**

A. Risk of Bias

List and describe predictors assessed, e.g. definition, timing and knowledge of other predictors: • Demographics factors: Age • Clinical history factors: Side-related symptoms: dominant arm involvement, Symptom duration • Patient reported measures: Pain (none) , Psychological symptoms (none), Function / Disability (none) • Clinical measures: Strength (none), ROM (none), Scapular movement and control (none) • Structural pathology: Imaging (none), Orthopaedic tests (none) All recorded at initiation of the study	
1. Were predictors defined and assessed in a similar way for all participants?	Yes
3. Were predictor assessments made without knowledge of outcome data?	Yes
4. Are all predictors available at the time the model is intended to be used?	Yes
5. Were all relevant predictors analysed?	Yes
Risk of bias introduced by predictors or their assessment	RISK: Low

B. Applicability

Concern that the definition, assessment or timing of assessment of predictors in the model do not match the review question	CONCERN: Low

DOMAIN 3: **Outcome**

A. Risk of Bias

Describe the outcome and how it was defined and determined: VAS pain score; upon completion of treatment	
1. Was a pre-specified outcome definition used?	Yes
2. Were predictors excluded from the outcome definition?	Yes
3. Was the outcome defined and determined in a similar way for all participants?	Yes
5. Was the outcome determined without knowledge of predictor information?	Yes
Risk of bias introduced by the outcome or its determination	RISK: Low

B. Applicability

At what time point was the outcome determined: 3 weeks; upon completion of treatment If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome: n/a	
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: Low

DOMAIN 4: **Sample size and participant flow**

A. Risk of Bias

Describe numbers of participants and events per predictor: • n=14; 4.7 Describe the time interval between predictor assessment and outcome determination: • 3 weeks Describe any participants who were excluded from the model: • No data Describe missing data on predictors and outcomes as well as methods used for missing data: • No data	
1. Were there a reasonable number of outcome events? Less than half of the recommended (10)	No
2. Was the time interval between predictor assessment and outcome determination appropriate?	Yes
3. Were all enrolled participants included in the analysis?	NI
4. Were participants with missing data handled appropriately?	NI
Risk of bias introduced by sample size or participant flow	RISK: High

DOMAIN 5: **Analysis**

A. Risk of Bias

Describe how the model was developed (predictor selection, fitting and optimism, risk groups, model performance): • No attempt made to trim the number of candidate variables • No data on model construction Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): • n/a Describe the performance measures of the model, e.g. calibration, discrimination, (re)classification, net benefit: • n/a	
1. Were non-binary predictors handled appropriately?	NI
2. Was selection of predictors based on univariable analysis avoided?	NI
3. Was model overfitting (optimism in model performance) accounted for, e.g. using bootstrapping or shrinkage techniques?	No
4. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	No
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	No
6. Were relevant model performance measures evaluated, e.g. calibration, discrimination, (re)classification and net benefit?	No
Risk of bias introduced by the analysis	RISK: High

Step 4: Overall judgement

Overall judgement about risk of bias and applicability of the prediction model evaluation	
Overall judgement of risk of bias	RISK: High
Overall judgement of applicability	CONCERN: High

Step 5: Usability of the model

Assess the usability of the model	
Is the model presented with sufficient detail to be used in the intended context and target population?	RATING: No

Deutscher et al. (2009)

Step 3: Assess risk of bias and applicability

DOMAIN 1: **Participant selection**

A. Risk of Bias

Describe the sources of data and criteria for participant selection: • Fifty-four community-based outpatient physical therapy clinics in Israel • Non-specific shoulder pain.	
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Yes
2. Were all inclusions and exclusions of participants appropriate? No details provided	NI
3. Were participants enrolled at a similar state of health, or were predictors considered to account for any dissimilarities?	Unclear
Risk of bias introduced by selection of participants	RISK: High

B. Applicability

Describe included participants and setting: n=5000 Fifty-four community-based outpatient physical therapy clinics in Israel	
Concern that the included participants and setting do not match the review question	CONCERN: High

DOMAIN 2: **Predictors**

A. Risk of Bias

List and describe predictors assessed, e.g. definition, timing and knowledge of other predictors: • Demographics factors: Age, gender, smoker status, physical activity levels • Clinical history factors: Symptom duration: chronic symptoms, Taking antidepressant medication, Taking pain medication, Co-morbidities: number of co-morbidities • Patient reported measures: Pain (none), Psychological symptoms (none), Function / Disability (CAT) • Clinical measures: Strength (none), ROM (none), Scapular movement and control (none) • Structural pathology: Imaging (none), Orthopaedic tests (none) All recorded at initiation of the study	
1. Were predictors defined and assessed in a similar way for all participants?	Yes
3. Were predictor assessments made without knowledge of outcome data?	Yes
4. Are all predictors available at the time the model is intended to be used?	Yes
5. Were all relevant predictors analysed?	Yes

Risk of bias introduced by predictors or their assessment	RISK: Low

B. Applicability

Concern that the definition, assessment or timing of assessment of predictors in the model do not match the review question	CONCERN: Low

DOMAIN 3: **Outcome**

A. Risk of Bias

Describe the outcome and how it was defined and determined: Functional status via absolute CAT score at discharge; upon completion of treatment	
1. Was a pre-specified outcome definition used?	Yes
2. Were predictors excluded from the outcome definition? CAT score was both a predictor and a determinant of outcome	No
3. Was the outcome defined and determined in a similar way for all participants?	Yes
5. Was the outcome determined without knowledge of predictor information?	Yes
Risk of bias introduced by the outcome or its determination	RISK: High

B. Applicability

At what time point was the outcome determined: Upon completion of treatment If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome: n/a	
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: Low

DOMAIN 4: **Sample size and participant flow**

A. Risk of Bias

Describe numbers of participants and events per predictor: • n=5000; 238 Describe the time interval between predictor assessment and outcome determination: • n/a Describe any participants who were excluded from the model: • No data Describe missing data on predictors and outcomes as well as methods used for missing data: • No data	
1. Were there a reasonable number of outcome events?	Yes

2. Was the time interval between predictor assessment and outcome determination appropriate?	NI
3. Were all enrolled participants included in the analysis?	NI
4. Were participants with missing data handled appropriately?	NI
Risk of bias introduced by sample size or participant flow	RISK: High

DOMAIN 5: **Analysis**

A. Risk of Bias

Describe how the model was developed (predictor selection, fitting and optimism, risk groups, model performance): • Reduced number of candidate variables by collapsing of variables within a category if distributions were too low in any one variable; the variable with the greatest clinical relevance was retained where collinearity amongst independent variables was identified. • Most parsimonious model via forward selection and backward elimination Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): • n/a Describe the performance measures of the model, e.g. calibration, discrimination, (re)classification, net benefit: • n/a	
1. Were non-binary predictors handled appropriately?	Yes
2. Was selection of predictors based on univariable analysis avoided?	Yes
3. Was model overfitting (optimism in model performance) accounted for, e.g. using bootstrapping or shrinkage techniques?	No
4. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	No
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	NI
6. Were relevant model performance measures evaluated, e.g. calibration, discrimination, (re)classification and net benefit?	No
Risk of bias introduced by the analysis	RISK: High

Step 4: Overall judgement

Overall judgement about risk of bias and applicability of the prediction model evaluation

Overall judgement of risk of bias	RISK: High
Overall judgement of applicability	CONCERN: High

Step 5: Usability of the model

Assess the usability of the model	
Is the model presented with sufficient detail to be used in the intended context and target population?	RATING: No

Engebretsen et al. (2010)

Step 3: Assess risk of bias and applicability

DOMAIN 1: **Participant selection**

A. Risk of Bias

Describe the sources of data and criteria for participant selection: • Recruited from orthopaedic clinics and university hospital outpatient departments, Norway • Patients between 18 and 70 years old with subacromial shoulder pain for at least 3 months. The inclusion criteria were: dysfunction or pain on abduction; a normal passive glenohumeral range of motion; pain on two of three isometric tests (abduction at 0° or 30°, external or internal rotation); and a positive impingement sign. The exclusion criteria were: bilateral shoulder pain, previous surgery on the affected shoulder, instability, referred pain from neck, rheumatoid arthritis, clinical and radiological signs of glenohumeral- or acromioclavicular arthritis, serious somatic or psychiatric disorder or inability to understand Norwegian.	
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Yes
2. Were all inclusions and exclusions of participants appropriate?	Yes
3. Were participants enrolled at a similar state of health, or were predictors considered to account for any dissimilarities?	Unclear
Risk of bias introduced by selection of participants	RISK: Low

B. Applicability

Describe included participants and setting: n=104 Recruited from orthopaedic clinics and university hospital outpatient departments, Norway	
Concern that the included participants and setting do not match the review question	CONCERN: Low

DOMAIN 2: **Predictors**

A. Risk of Bias

List and describe predictors assessed, e.g. definition, timing and knowledge of other predictors: • Demographics factors: Age, gender, Work status, Occupation: frequency of working above shoulder height; frequency of carrying heavy loads at work, educational attainment

• Clinical history factors: Recurrent shoulder pain: previous shoulder pain, Symptom duration, Associated neck pain, Side-related symptoms: dominant arm involvement, Previously receiving treatment: previous physiotherapy, Taking medication • Patient reported measures: Pain (pain intensity), Psychological symptoms (General health status, Emotional distress: Hopkins Symptom checklist, self-efficacy for pain), Function / Disability (SPADI) • Clinical measures: Strength (none), ROM (AROM – HBB, flexion on the affected side), Scapular movement and control (none) • Structural pathology: Imaging (none), Orthopaedic tests (none) All recorded at initiation of the study	
1. Were predictors defined and assessed in a similar way for all participants?	Yes
3. Were predictor assessments made without knowledge of outcome data?	Yes
4. Are all predictors available at the time the model is intended to be used?	Yes
5. Were all relevant predictors analysed?	Yes
Risk of bias introduced by predictors or their assessment	RISK: Low

B. Applicability

Concern that the definition, assessment or timing of assessment of predictors in the model do not match the review question	CONCERN: Low

DOMAIN 3: **Outcome**

A. Risk of Bias

Describe the outcome and how it was defined and determined: Pain and disability via absolute SPADI score; 'not working' status at 12 months post baseline	
1. Was a pre-specified outcome definition used?	Yes
2. Were predictors excluded from the outcome definition? SPADI was both a predictor and a determinant of outcome	No
3. Was the outcome defined and determined in a similar way for all participants?	Yes
5. Was the outcome determined without knowledge of predictor information?	Yes
Risk of bias introduced by the outcome or its determination	RISK: High

B. Applicability

At what time point was the outcome determined: 12 months post baseline If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome: n/a	
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: Low

DOMAIN 4: **Sample size and participant flow**

A. Risk of Bias

Describe numbers of participants and events per predictor: • n=1.4; 5.5 Describe the time interval between predictor assessment and outcome determination: • 12 months Describe any participants who were excluded from the model: • 10%; analysis of differences between groups = missing subjects older and with a higher level of functional impairment at baseline compared to those for whom follow up at 1 year was possible Describe missing data on predictors and outcomes as well as methods used for missing data: • No data	
1. Were there a reasonable number of outcome events? Almost half of the recommended 10	No
2. Was the time interval between predictor assessment and outcome determination appropriate?	Yes
3. Were all enrolled participants included in the analysis?	Yes
4. Were participants with missing data handled appropriately?	NI
Risk of bias introduced by sample size or participant flow	RISK: High

DOMAIN 5: **Analysis**

A. Risk of Bias

Describe how the model was developed (predictor selection, fitting and optimism, risk groups, model performance): • Reduced number of candidate variables via statistical methods – looking at univariate relationships between variables; the variable with the greatest clinical relevance was retained where collinearity amongst independent variables was identified. • Most parsimonious model via variables with $p \geq 0.05$ removed Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): • n/a Describe the performance measures of the model, e.g. calibration, discrimination, (re)classification, net benefit: • n/a	
1. Were non-binary predictors handled appropriately?	Yes
2. Was selection of predictors based on univariable analysis avoided?	No
3. Was model overfitting (optimism in model performance) accounted for, e.g. using bootstrapping or shrinkage techniques?	No
4. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	No
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	Yes
6. Were relevant model performance measures evaluated, e.g. calibration, discrimination, (re)classification and net benefit?	No
Risk of bias introduced by the analysis	RISK: High

Step 4: Overall judgement

Overall judgement about risk of bias and applicability of the prediction model evaluation	
Overall judgement of risk of bias	RISK: High
Overall judgement of applicability	CONCERN: High

Step 5: Usability of the model

Assess the usability of the model	
Is the model presented with sufficient detail to be used in the intended context and target population?	RATING: No

Hung et al (2010)

Step 3: Assess risk of bias and applicability

DOMAIN 1: **Participant selection**

A. Risk of Bias

Describe the sources of data and criteria for participant selection: • Recruited patients with SAIS from the orthopedics clinic in National Taiwan University Hospital and also through general announcements in local Internet media • Subjects had to demonstrate at least 3 of the following: (1) a positive Neer impingement test, (2) a positive Hawkins impingement test, (3) a painful arc, (4 pain with isometric resisted abduction, (5) pain with palpation of the rotator cuff tendons, and (6) pain with active shoulder elevation. Subjects were excluded if they demonstrated signs of a complete rotator cuff tear or acute inflammation.	
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Yes
2. Were all inclusions and exclusions of participants appropriate?	Yes
3. Were participants enrolled at a similar state of health, or were predictors considered to account for any dissimilarities?	Unclear
Risk of bias introduced by selection of participants	RISK: Low

B. Applicability

Describe included participants and setting: n=33 From orthopedics clinic in National Taiwan University Hospital and also through general announcements in local Internet media	
Concern that the included participants and setting do not match the review question	CONCERN: Low

DOMAIN 2: **Predictors**

A. Risk of Bias

List and describe predictors assessed, e.g. definition, timing and knowledge of other predictors: • Demographics factors: Age • Clinical history factors: None • Patient reported measures: Pain (none), Psychological symptoms (none), Function / Disability (FLEX-SF) • Clinical measures: Strength (scapular protractor force, ER, IR and abduction), ROM (PROM IR, ER; posterior shoulder tightness (PST)), Scapular movement and control (Scapular kinematics (UR, IR and scapular tilt) at 30°, 60°, 90°, 120° elevation during both the ascending and descending phases in loaded and unloaded conditions) • Structural pathology: Imaging (none), Orthopaedic tests (none) All recorded at initiation of the study	
1. Were predictors defined and assessed in a similar way for all participants?	Yes
3. Were predictor assessments made without knowledge of outcome data?	Yes
4. Are all predictors available at the time the model is intended to be used?	No FASTRAK motion analysis system prohibitive in clinical setting
5. Were all relevant predictors analysed?	Yes
Risk of bias introduced by predictors or their assessment	RISK: High

B. Applicability

Concern that the definition, assessment or timing of assessment of predictors in the model do not match the review question	CONCERN: High 3 dimensional movement analysis system prohibitive in clinical setting

DOMAIN 3: **Outcome**

A. Risk of Bias

Describe the outcome and how it was defined and determined: 'Improvement' via GROC score => 4 at discharge; upon completion of treatment	
1. Was a pre-specified outcome definition used?	Yes
2. Were predictors excluded from the outcome definition?	Yes
3. Was the outcome defined and determined in a similar way for all participants?	Yes
5. Was the outcome determined without knowledge of predictor information?	Yes
Risk of bias introduced by the outcome or its determination	RISK: Low

B. Applicability

At what time point was the outcome determined: At end of 6 weeks of treatment If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome: n/a

Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: Low

DOMAIN 4: **Sample size and participant flow**

A. Risk of Bias

Describe numbers of participants and events per predictor: • n=33; 0.5 Describe the time interval between predictor assessment and outcome determination: • 6 weeks Describe any participants who were excluded from the model: • 1; no data Describe missing data on predictors and outcomes as well as methods used for missing data: • n/a	
1. Were there a reasonable number of outcome events? Substantially below the 10 threshold	No
2. Was the time interval between predictor assessment and outcome determination appropriate?	Yes
3. Were all enrolled participants included in the analysis?	No
4. Were participants with missing data handled appropriately?	n/a
Risk of bias introduced by sample size or participant flow	RISK: High

DOMAIN 5: **Analysis**

A. Risk of Bias

Describe how the model was developed (predictor selection, fitting and optimism, risk groups, model performance): • Reduced number of candidate variables via statistical methods – looking at univariate relationships between variables Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): • n/a Describe the performance measures of the model, e.g. calibration, discrimination, (re)classification, net benefit: • n/a	
1. Were non-binary predictors handled appropriately?	Yes
2. Was selection of predictors based on univariable analysis avoided?	No
3. Was model overfitting (optimism in model performance) accounted for, e.g. using bootstrapping or shrinkage techniques?	No
4. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	No
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	No

6. Were relevant model performance measures evaluated, e.g. calibration, discrimination, (re)classification and net benefit?	No
Risk of bias introduced by the analysis	RISK: High

Step 4: Overall judgement

Overall judgement about risk of bias and applicability of the prediction model evaluation	
Overall judgement of risk of bias	RISK: High
Overall judgement of applicability	CONCERN: High

Step 5: Usability of the model

Assess the usability of the model	
Is the model presented with sufficient detail to be used in the intended context and target population?	RATING: No

Kennedy et al. (2006)

Step 3: Assess risk of bias and applicability

DOMAIN 1: **Participant selection**

A. Risk of Bias

Describe the sources of data and criteria for participant selection: • Patients beginning treatment at a large number of different physiotherapy practices; Canada • Non-specific shoulder pain Soft tissue shoulder disorders: Shoulder complaints were defined as any condition of pain or discomfort, including instances where there had been surgical treatment of the soft tissue shoulder disorder (eg, rotator cuff repair). 8% post-surgery. Patients were excluded from the study if they: (1) had fractures or dislocations associated with soft tissue pain, (2) received physical therapy for only one visit (eg, referred for equipment or single education session), or (3) were unable to read and write English and thus could not complete the questionnaire package independently.	
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Yes
2. Were all inclusions and exclusions of participants appropriate?	Yes
3. Were participants enrolled at a similar state of health, or were predictors considered to account for any dissimilarities?	Unclear
Risk of bias introduced by selection of participants	RISK: Low

B. Applicability

Describe included participants and setting: n=361 Patients beginning treatment at a large number of different physiotherapy practices; Canada	
Concern that the included participants and setting do not match the review question	CONCERN: High

DOMAIN 2: **Predictors**

A. Risk of Bias

List and describe predictors assessed, e.g. definition, timing and knowledge of other predictors: • Demographics factors: Age, gender, Work status, workers compensation claim • Clinical history factors: Symptom duration: shorter duration in symptoms, Post-surgical case, Recurrent shoulder pain, Reason for symptom onset, Co-morbidities: number of co-morbidities and number that limit activity, Taking pain medication • Patient reported measures: Pain (pain intensity), Psychological symptoms (Patient expectation, Mental Component Score (MCS) from the 36-Item Short-Form Health Survey (SF-36)), Function / Disability (DASH, Physical Component Score (PCS) from the 36-Item Short-Form Health Survey (SF-36), Global rating of disability score) • Clinical measures: Strength (muscle wasting and muscle weakness), ROM (active and passive ROM), Scapular movement and control (none) • Structural pathology: Imaging (none), Orthopaedic tests (none) All recorded at initiation of the study	
1. Were predictors defined and assessed in a similar way for all participants?	Yes
3. Were predictor assessments made without knowledge of outcome data?	Yes
4. Are all predictors available at the time the model is intended to be used?	Yes
5. Were all relevant predictors analysed?	Yes
Risk of bias introduced by predictors or their assessment	RISK: Low

B. Applicability

Concern that the definition, assessment or timing of assessment of predictors in the model do not match the review question	CONCERN: Low

DOMAIN 3: **Outcome**

A. Risk of Bias

Describe the outcome and how it was defined and determined: Disability via absolute DASH score at discharge / 12 weeks; change in disability via change from DASH score at baseline to discharge / 12 weeks	
1. Was a pre-specified outcome definition used?	Yes
2. Were predictors excluded from the outcome definition? DASH both predictor and determinant of outcome	No
3. Was the outcome defined and determined in a similar way for all participants?	Yes
5. Was the outcome determined without knowledge of predictor information?	Yes
Risk of bias introduced by the outcome or its determination	RISK: High

B. Applicability

At what time point was the outcome determined: Discharge or 12 weeks post baseline If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome: n/a	
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: Low

DOMAIN 4: **Sample size and participant flow**

A. Risk of Bias

Describe numbers of participants and events per predictor: • n=361; 10 Describe the time interval between predictor assessment and outcome determination: • Discharge or 12 weeks post baseline Describe any participants who were excluded from the model: • no data Describe missing data on predictors and outcomes as well as methods used for missing data: • n/a	
1. Were there a reasonable number of outcome events?	Yes
2. Was the time interval between predictor assessment and outcome determination appropriate?	Yes
3. Were all enrolled participants included in the analysis?	No
4. Were participants with missing data handled appropriately?	n/a
Risk of bias introduced by sample size or participant flow	RISK: High

DOMAIN 5: **Analysis**

A. Risk of Bias

Describe how the model was developed (predictor selection, fitting and optimism, risk groups, model performance): • Reduced number of candidate variables via statistical methods – looking at univariate relationships between variables; also collapsing of variables within a category if distributions were too low in any one variable; the variable with the greatest clinical relevance was retained where collinearity amongst independent variables was identified Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): • n/a Describe the performance measures of the model, e.g. calibration, discrimination, (re)classification, net benefit: • n/a	
1. Were non-binary predictors handled appropriately?	Yes
2. Was selection of predictors based on univariable analysis avoided?	No

3. Was model overfitting (optimism in model performance) accounted for, e.g. using bootstrapping or shrinkage techniques?	No
4. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	No
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	Yes
6. Were relevant model performance measures evaluated, e.g. calibration, discrimination, (re)classification and net benefit?	No
Risk of bias introduced by the analysis	RISK: High

Step 4: Overall judgement

Overall judgement about risk of bias and applicability of the prediction model evaluation	
Overall judgement of risk of bias	RISK: High
Overall judgement of applicability	CONCERN: High

Step 5: Usability of the model

Assess the usability of the model	
Is the model presented with sufficient detail to be used in the intended context and target population?	RATING: No

Kromer et al (2014)

Step 3: Assess risk of bias and applicability

DOMAIN 1: **Participant selection**

A. Risk of Bias

Describe the sources of data and criteria for participant selection:

- Germany, outpatient. Presentation to a physical therapist following referral by general practitioner or orthopedic surgeon (duration of symptoms >=4 wk)
  Outpatient physiotherapy clinics, GPs and orthopaedic surgeons
- Inclusion criteria set for this trial were: (1) age between 18 and 75 years; (2) symptoms for at least 4 weeks; (3) main complaints in the glenohumeral joint region or the proximal segments of the arm; (4) presence of one of the following signs indicating SPS: Neer impingement sign, Hawkins- Kennedy impingement test, or painful arc with active abduction or flexion; and (5) pain during one of the following resistance tests: external rotation, internal rotation, abduction, or flexion.
  Exclusion criteria were: (1) average 24-hour pain of 8/10 or more on a visual numeric rating scale (VNRS); (2) primary scapulothoracic dysfunction due to paresis; (3) diagnosed instability or previous history of dislocation; (4) frozen shoulder; (5) more than one-third restriction of elevation compared with the unaffected side; (6) substantial shoulder weakness or loss of active shoulder function; (7) shoulder surgery on the involved side in the previous 12 months; (8) reproduction of symptoms with active or passive cervical movements; (9) neurological involvement with sensory or muscular deficits; (10) inflammatory joint disease (eg, rheumatoid arthritis); (11) diabetes mellitus; (12) intake of psychotherapeutic drugs; (13) compensation claims; and (14) inability to understand written or spoken German.

1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Yes
2. Were all inclusions and exclusions of participants appropriate?	Yes
3. Were participants enrolled at a similar state of health, or were predictors considered to account for any dissimilarities?	Yes
Risk of bias introduced by selection of participants	RISK: Low

B. Applicability

Describe included participants and setting: n=90 Germany, outpatient. Presentation to a physical therapist following referral by general practitioner or orthopedic surgeon (duration of symptoms >=4 wk) Outpatient physiotherapy clinics, GPs and orthopaedic surgeons	
Concern that the included participants and setting do not match the review question	CONCERN: Low

DOMAIN 2: **Predictors**

A. Risk of Bias

List and describe predictors assessed, e.g. definition, timing and knowledge of other predictors: • Demographics factors: Age and gender were controlled for in the regression model • Clinical history factors: Symptom duration • Patient reported measures: Pain (pain intensity), Psychological symptoms (Fear Avoidance Beliefs Questionnaire (physical activity subscale), Pain catastrophising scale), Function / Disability (Function / disability as measured by the function element of the SPADI) • Clinical measures: Strength (none), ROM (none), Scapular movement and control (none) • Structural pathology: Imaging (none), Orthopaedic tests (none) All recorded at initiation of the study	
1. Were predictors defined and assessed in a similar way for all participants?	Yes
3. Were predictor assessments made without knowledge of outcome data?	Yes
4. Are all predictors available at the time the model is intended to be used?	Yes
5. Were all relevant predictors analysed?	Yes
Risk of bias introduced by predictors or their assessment	RISK: Low

B. Applicability

Concern that the definition, assessment or timing of assessment of predictors in the model do not match the review question	CONCERN: Low

DOMAIN 3: **Outcome**

A. Risk of Bias

Describe the outcome and how it was defined and determined: Change in functional status via change in functional subscale of SPADI (SPADI-F) between baseline and 3 months post baseline

1. Was a pre-specified outcome definition used?	Yes
2. Were predictors excluded from the outcome definition? SPADI-F also one of the predictors	No
3. Was the outcome defined and determined in a similar way for all participants?	Yes
5. Was the outcome determined without knowledge of predictor information?	Yes
Risk of bias introduced by the outcome or its determination	RISK: High

B. Applicability

At what time point was the outcome determined: 3 months post baseline If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome: n/a	
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: Low

DOMAIN 4: **Sample size and participant flow**

A. Risk of Bias

Describe numbers of participants and events per predictor: • n=90; 18 Describe the time interval between predictor assessment and outcome determination: • 3 months Describe any participants who were excluded from the model: • 2; no data Describe missing data on predictors and outcomes as well as methods used for missing data: • n/a	
1. Were there a reasonable number of outcome events?	Yes
2. Was the time interval between predictor assessment and outcome determination appropriate?	Yes
3. Were all enrolled participants included in the analysis?	No
4. Were participants with missing data handled appropriately?	n/a
Risk of bias introduced by sample size or participant flow	RISK: High

DOMAIN 5: **Analysis**

A. Risk of Bias

Describe how the model was developed (predictor selection, fitting and optimism, risk groups, model performance):

• Reduced number of candidate variables via variable with the greatest clinical relevance was retained where collinearity amongst independent variables was identified Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): • n/a Describe the performance measures of the model, e.g. calibration, discrimination, (re)classification, net benefit: • n/a	
1. Were non-binary predictors handled appropriately?	Yes
2. Was selection of predictors based on univariable analysis avoided?	Yes
3. Was model overfitting (optimism in model performance) accounted for, e.g. using bootstrapping or shrinkage techniques?	No
4. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	No
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	No
6. Were relevant model performance measures evaluated, e.g. calibration, discrimination, (re)classification and net benefit?	No
Risk of bias introduced by the analysis	RISK: High

Step 4: Overall judgement

Overall judgement about risk of bias and applicability of the prediction model evaluation	
Overall judgement of risk of bias	RISK: High
Overall judgement of applicability	CONCERN: Low

Step 5: Usability of the model

Assess the usability of the model	
Is the model presented with sufficient detail to be used in the intended context and target population?	RATING: No

Mintken et al. (2010)

Step 3: Assess risk of bias and applicability

DOMAIN 1: **Participant selection**

A. Risk of Bias

Describe the sources of data and criteria for participant selection: • Outpatient physiotherapy clinics • Non-specific shoulder pain Inclusion criteria: between ages of 18 and 65 years, with a primary report of shoulder pain and a baseline Shoulder Pain and Disability Index (SPADI) score of 20% or greater.

Exclusion criteria: any medical “red flags” suggestive of a non-musculoskeletal aetiology of symptoms, acute fractures in the shoulder region, acute severe trauma in the cervical or thoracic region in the previous 6 weeks, a diagnosis of cervical spinal stenosis or bilateral upper-extremity symptoms, osteoporosis, prior surgery to the cervical or thoracic region, evidence of central nervous system involvement, insufficient English language skills to complete the questionnaires, or signs consistent with nerve root compression (defined as impairment in at least 2 of the following: myotomal strength, sensation, or reflexes). “Red flags” were ruled out by a combination of a medical screening questionnaire, a neurological examination, and a patient history.	
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Yes
2. Were all inclusions and exclusions of participants appropriate?	Yes
3. Were participants enrolled at a similar state of health, or were predictors considered to account for any dissimilarities?	Uncertain
Risk of bias introduced by selection of participants	RISK: Low

B. Applicability

Describe included participants and setting: n=80 Outpatient physiotherapy clinics	
Concern that the included participants and setting do not match the review question	CONCERN: Low

DOMAIN 2: **Predictors**

A. Risk of Bias

List and describe predictors assessed, e.g. definition, timing and knowledge of other predictors: • Demographics factors: Age, gender, work status • Clinical history factors: Symptom duration, Taking pain medication, Reason for symptom onset, Recurrent shoulder pain: number of previous episodes, Previously receiving treatment: treatment for previous episodes, past medical history • Patient reported measures: Pain (pain intensity), Psychological symptoms (Patient expectation, Fear Avoidance Beliefs Questionnaire, Tampa Scale for Kinesiophobia), Function / Disability (SPADI) • Clinical measures: Strength (Serratus Anterior, Middle trapezius, Lower trapezius, Rhomboid, Deltoid, External and internal shoulder rotator muscle strength), ROM (pain-free shoulder flexion, passive shoulder IR at 90° abduction, Passive shoulder abduction, Passive shoulder ER at 90° abduction, Battery of 3 functional tests), Scapular movement and control (Lateral slide test and scapula index, Qualitative assessment of scapular function) • Structural pathology: Imaging (none), Orthopaedic tests (13 orthopaedic tests including the Hawkins-Kennedy impingement test, the empty can and full can test and the drop sign Neer's sign) All recorded at initiation of the study	
1. Were predictors defined and assessed in a similar way for all participants?	Yes
3. Were predictor assessments made without knowledge of outcome data?	Yes
4. Are all predictors available at the time the model is intended to be used?	Yes
5. Were all relevant predictors analysed?	Yes
Risk of bias introduced by predictors or their assessment	RISK: Low

B. Applicability

Concern that the definition, assessment or timing of assessment of predictors in the model do not match the review question	CONCERN: Low

DOMAIN 3: **Outcome**

A. Risk of Bias

Describe the outcome and how it was defined and determined: 'Improvement' via GROC score => 4 at discharge; upon completion of treatment At the beginning of the second session, the participants completed the GROC and the other outcome measures. If their score on the GROC did not exceed the +4 cutoff at the second session, they received the same intervention program again and were scheduled for a follow-up within 2 to 4 days. Participants again completed the GROC along with the other outcome measures. If they scored +4 or better on the GROC, they were categorized as having a successful outcome; if they scored below +4, they were categorized as not having a successful outcome. At this point, their participation in the study was complete, and the therapist could administer further treatment as needed. F/up was at 2nd or 3rd apt – over a period of a few days	
1. Was a pre-specified outcome definition used?	Yes
2. Were predictors excluded from the outcome definition?	Yes
3. Was the outcome defined and determined in a similar way for all participants?	Yes
5. Was the outcome determined without knowledge of predictor information?	Yes
Risk of bias introduced by the outcome or its determination	RISK: Low

B. Applicability

At what time point was the outcome determined: At completion of 2nd or 3rd appointment – over a period of a few days If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome: n/a	
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: High

DOMAIN 4: **Sample size and participant flow**

A. Risk of Bias

Describe numbers of participants and events per predictor: • n=80; 6 Describe the time interval between predictor assessment and outcome determination: • a few days Describe any participants who were excluded from the model: • 1; no data Describe missing data on predictors and outcomes as well as methods used for missing data: • n/a	
1. Were there a reasonable number of outcome events?	No

Just over half of the recommended number	
2. Was the time interval between predictor assessment and outcome determination appropriate? Very limited clinical relevance of a few days post commencement of treatment	No
3. Were all enrolled participants included in the analysis?	No
4. Were participants with missing data handled appropriately?	n/a
Risk of bias introduced by sample size or participant flow	RISK: High

DOMAIN 5: **Analysis**

A. Risk of Bias

Describe how the model was developed (predictor selection, fitting and optimism, risk groups, model performance): • Reduced number of candidate variables via statistical methods – looking at univariate relationships between variables • Variables with a significance level of P<.10 were retained as potential prognostic variables. For continuous variables with a significant univariate relationship, sensitivity and specificity values were calculated for all possible cutoff points and then plotted as a receiver operator characteristic (ROC) curve Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): • n/a Describe the performance measures of the model, e.g. calibration, discrimination, (re)classification, net benefit: • n/a	
1. Were non-binary predictors handled appropriately?	Yes
2. Was selection of predictors based on univariable analysis avoided?	No
3. Was model overfitting (optimism in model performance) accounted for, e.g. using bootstrapping or shrinkage techniques?	No
4. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	No
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	Yes
6. Were relevant model performance measures evaluated, e.g. calibration, discrimination, (re)classification and net benefit?	No
Risk of bias introduced by the analysis	RISK: High

Step 4: Overall judgement

Overall judgement about risk of bias and applicability of the prediction model evaluation	
Overall judgement of risk of bias	RISK: High
Overall judgement of applicability	CONCERN: High

Step 5: Usability of the model

Assess the usability of the model	
Is the model presented with sufficient detail to be used in the intended context and target population?	RATING: No

Morrison et al. (1997)

Step 3: Assess risk of bias and applicability

DOMAIN 1: **Participant selection**

A. Risk of Bias

Describe the sources of data and criteria for participant selection: • Investigation performed at the Southern California Centre for Sports Medicine, Long Beach • 'Subacromial impingement' Diagnosis made on the basis of a positive impingement sign and the absence of other abnormalities of the shoulder, such as full thickness tears of the rotator cuff, osteoarthrosis of the acromioclavicular joint, instability of the glenohumeral joint, or adhesive capsulitis. Diagnosis made on the basis of a history, a clinical examination, and a positive Neer impingement sign. Patients who had concomitant adhesive capsulitis, cervical radiculopathy, or suprascapular nerve palsy were excluded, as were patients who had major weakness on testing of the rotator cuff or a full-thickness tear of the rotator cuff on magnetic resonance imaging..	
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Yes
2. Were all inclusions and exclusions of participants appropriate?	Yes
3. Were participants enrolled at a similar state of health, or were predictors considered to account for any dissimilarities?	Uncertain
Risk of bias introduced by selection of participants	RISK: Low

B. Applicability

Describe included participants and setting: n=616 Investigation performed at the Southern California Centre for Sports Medicine, Long Beach	
Concern that the included participants and setting do not match the review question	CONCERN: Low

DOMAIN 2: **Predictors**

A. Risk of Bias

List and describe predictors assessed, e.g. definition, timing and knowledge of other predictors: • Demographics factors: Age, gender • Clinical history factors: Symptom duration, Side-related symptoms: dominant arm involvement • Patient reported measures: Pain (none), Psychological symptoms (none), Function / Disability (none) • Clinical measures: Strength (none), ROM (none), Scapular movement and control (none)

• Structural pathology: Imaging (Acromion morphology type 1 on X-ray), Orthopaedic tests (none) All recorded at initiation of the study	
1. Were predictors defined and assessed in a similar way for all participants?	Yes
3. Were predictor assessments made without knowledge of outcome data?	Yes
4. Are all predictors available at the time the model is intended to be used? Routine x-ray is not performed	No
5. Were all relevant predictors analysed?	Yes
Risk of bias introduced by predictors or their assessment	RISK: High

B. Applicability

Concern that the definition, assessment or timing of assessment of predictors in the model do not match the review question	CONCERN: Low

DOMAIN 3: **Outcome**

A. Risk of Bias

Describe the outcome and how it was defined and determined: Pain, function, active ROM, strength and overall satisfaction via absolute UCLA score at average of 27 months	
1. Was a pre-specified outcome definition used?	Yes
2. Were predictors excluded from the outcome definition?	Yes
3. Was the outcome defined and determined in a similar way for all participants?	Yes
5. Was the outcome determined without knowledge of predictor information?	Yes
Risk of bias introduced by the outcome or its determination	RISK: Low

B. Applicability

At what time point was the outcome determined: At average of 27 months post baseline If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome: n/a	
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: Low

DOMAIN 4: **Sample size and participant flow**

A. Risk of Bias

Describe numbers of participants and events per predictor: • n=616; 123 Describe the time interval between predictor assessment and outcome determination: • Average of 27 months

Describe any participants who were excluded from the model: • 8%; no data Describe missing data on predictors and outcomes as well as methods used for missing data: • n/a	
1. Were there a reasonable number of outcome events?	Yes
2. Was the time interval between predictor assessment and outcome determination appropriate?	Yes
3. Were all enrolled participants included in the analysis?	No
4. Were participants with missing data handled appropriately?	n/a
Risk of bias introduced by sample size or participant flow	RISK: High

DOMAIN 5: **Analysis**

A. Risk of Bias

Describe how the model was developed (predictor selection, fitting and optimism, risk groups, model performance): • No attempt made to trim the number of candidate variables • Chi-square Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): • n/a Describe the performance measures of the model, e.g. calibration, discrimination, (re)classification, net benefit: • n/a	
1. Were non-binary predictors handled appropriately?	Yes
2. Was selection of predictors based on univariable analysis avoided?	No
3. Was model overfitting (optimism in model performance) accounted for, e.g. using bootstrapping or shrinkage techniques?	No
4. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	No
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	No
6. Were relevant model performance measures evaluated, e.g. calibration, discrimination, (re)classification and net benefit?	No
Risk of bias introduced by the analysis	RISK: High

Step 4: Overall judgement

Overall judgement about risk of bias and applicability of the prediction model evaluation	
Overall judgement of risk of bias	RISK: High
Overall judgement of applicability	CONCERN: Low

Step 5: Usability of the model

Assess the usability of the model	
Is the model presented with sufficient detail to be used in the intended context and target population?	RATING: No

Ogon et al. (2009)

Step 3: Assess risk of bias and applicability

DOMAIN 1: **Participant selection**

A. Risk of Bias

Describe the sources of data and criteria for participant selection: • All patients had previously received nonoperative treatment from general practitioners, rheumatologists, or orthopedic surgeons, which included physical therapy, manual therapy, electrotherapy, iontophoresis, systemic use of analgesics and nonsteroidal antiinflammatory drugs (NSAIDs), and up to 3 subacromial injections of corticosteroids. The patients were then referred to the Orthopedic Outpatient Clinic because of the persistence of clinically symptomatic calcific tendinitis of the shoulder • 'Calcific tendonitis' Inclusion criteria were the presence of radiographically and sonographically proven calcific deposits in a rotator cuff tendon and the presence of clinically symptomatic calcific tendinitis of the shoulder requiring continuation of treatment at the time of presentation at the institution. Exclusion criteria were previous surgical interventions, needling, application of ultrasound therapy, or extracorporeal shock wave therapy (ESWT), as well as the presence of rheumatoid arthritis or concomitant diseases of the affected shoulder.	
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Yes
2. Were all inclusions and exclusions of participants appropriate?	Yes
3. Were participants enrolled at a similar state of health, or were predictors considered to account for any dissimilarities?	Yes
Risk of bias introduced by selection of participants	RISK: Low

B. Applicability

Describe included participants and setting: n=420 Orthopaedic Outpatient Clinic	
Concern that the included participants and setting do not match the review question	CONCERN: High

DOMAIN 2: **Predictors**

A. Risk of Bias

List and describe predictors assessed, e.g. definition, timing and knowledge of other predictors: • Demographics factors: Age, gender, occupation • Clinical history factors: Periods of professional disability, Reason for symptom onset, Symptom duration, Side-related symptoms: dominant arm involvement, Past medical history • Patient reported measures: Pain (none), Psychological symptoms (none), Function / Disability (none) • Clinical measures: Strength (none), ROM (none), Scapular movement and control (none) • Structural pathology: Imaging (Calcific deposits: X-ray and sonographic), Orthopaedic tests (none) All recorded at initiation of the study	
1. Were predictors defined and assessed in a similar way for all participants?	Yes
3. Were predictor assessments made without knowledge of outcome data?	Yes
4. Are all predictors available at the time the model is intended to be used? Routine x-ray and ultrasound is not performed	No
5. Were all relevant predictors analysed?	Yes
Risk of bias introduced by predictors or their assessment	RISK: High

B. Applicability

Concern that the definition, assessment or timing of assessment of predictors in the model do not match the review question	CONCERN: High

DOMAIN 3: **Outcome**

A. Risk of Bias

Describe the outcome and how it was defined and determined: 'Success of non-operative therapy' via no progression to advanced therapeutic measures after a minimum of 6 months non-operative treatment (including minimum of 3 months treatment at the study location)	
1. Was a pre-specified outcome definition used?	Yes
2. Were predictors excluded from the outcome definition?	Yes
3. Was the outcome defined and determined in a similar way for all participants?	Yes
5. Was the outcome determined without knowledge of predictor information?	Yes
Risk of bias introduced by the outcome or its determination	RISK: Low

B. Applicability

At what time point was the outcome determined: After a minimum of 6 months non-operative treatment If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome: n/a	
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: High

DOMAIN 4: **Sample size and participant flow**

A. Risk of Bias

Describe numbers of participants and events per predictor: • n=420; 25 Describe the time interval between predictor assessment and outcome determination: • After a minimum of 6 months Describe any participants who were excluded from the model: • no data Describe missing data on predictors and outcomes as well as methods used for missing data: • n/a	
1. Were there a reasonable number of outcome events?	Yes
2. Was the time interval between predictor assessment and outcome determination appropriate?	Yes
3. Were all enrolled participants included in the analysis?	NI
4. Were participants with missing data handled appropriately?	n/a
Risk of bias introduced by sample size or participant flow	RISK: Uncertain

DOMAIN 5: **Analysis**

A. Risk of Bias

Describe how the model was developed (predictor selection, fitting and optimism, risk groups, model performance): • No attempt made to trim the number of candidate variables • Most parsimonious model via variables with $p \geq 0.05$ removed • Prognostic factors were determined at $P < 0.05$ by chi-square test Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): • n/a Describe the performance measures of the model, e.g. calibration, discrimination, (re)classification, net benefit: • n/a	
1. Were non-binary predictors handled appropriately?	Yes
2. Was selection of predictors based on univariable analysis avoided?	No
3. Was model overfitting (optimism in model performance) accounted for, e.g. using bootstrapping or shrinkage techniques?	No
4. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	No
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	No
6. Were relevant model performance measures evaluated, e.g. calibration, discrimination, (re)classification and net benefit?	No
Risk of bias introduced by the analysis	RISK: High

Step 4: Overall judgement

Overall judgement about risk of bias and applicability of the prediction model evaluation	
Overall judgement of risk of bias	RISK: High
Overall judgement of applicability	CONCERN: High

Step 5: Usability of the model

Assess the usability of the model	
Is the model presented with sufficient detail to be used in the intended context and target population?	RATING: No

Sindhu et al. (2012)

Step 3: Assess risk of bias and applicability

DOMAIN 1: **Participant selection**

A. Risk of Bias

Describe the sources of data and criteria for participant selection: • Retrospective cohort study Non-specific shoulder pain Clinical staff entered necessary medical information at intake, such as diagnosis codes based on the International Classification of Diseases, Ninth Revision (ICD-9).	
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Yes
2. Were all inclusions and exclusions of participants appropriate?	Yes
3. Were participants enrolled at a similar state of health, or were predictors considered to account for any dissimilarities?	Uncertain
Risk of bias introduced by selection of participants	RISK: Low

B. Applicability

Describe included participants and setting: n=3362 Outpatient rehabilitation clinics throughout the United States	
Concern that the included participants and setting do not match the review question	CONCERN: High

DOMAIN 2: **Predictors**

A. Risk of Bias

List and describe predictors assessed, e.g. definition, timing and knowledge of other predictors: • Demographics factors: None • Clinical history factors: None

• Patient reported measures: Pain (none), Psychological symptoms (Fear Avoidance Beliefs Questionnaire), Function / Disability (none) • Clinical measures: Strength (none), ROM (none), Scapular movement and control (none) • Structural pathology: Imaging (none), Orthopaedic tests (none) All recorded at initiation of the study	
1. Were predictors defined and assessed in a similar way for all participants?	Yes
3. Were predictor assessments made without knowledge of outcome data?	Yes
4. Are all predictors available at the time the model is intended to be used?	Yes
5. Were all relevant predictors analysed? All variables were controlled for except for the single variable of fear avoidance	No
Risk of bias introduced by predictors or their assessment	RISK: High

B. Applicability

Concern that the definition, assessment or timing of assessment of predictors in the model do not match the review question	CONCERN: High

DOMAIN 3: **Outcome**

A. Risk of Bias

Describe the outcome and how it was defined and determined: Functional change via change in CAT score between baseline and discharge; upon completion of treatment	
1. Was a pre-specified outcome definition used?	Yes
2. Were predictors excluded from the outcome definition?	Yes
3. Was the outcome defined and determined in a similar way for all participants?	Yes
5. Was the outcome determined without knowledge of predictor information?	Yes
Risk of bias introduced by the outcome or its determination	RISK: Low

B. Applicability

At what time point was the outcome determined: Upon completion of treatment If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome: n/a	
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: Low

DOMAIN 4: **Sample size and participant flow**

A. Risk of Bias

Describe numbers of participants and events per predictor: • n=3362; 3362

Describe the time interval between predictor assessment and outcome determination: • Upon completion of treatment Describe any participants who were excluded from the model: • Loss to follow up: 43%; analysis of differences between groups = identified differences between those available for and lost to follow up in terms of age, pain levels and function but provided no details on the direction of difference Describe missing data on predictors and outcomes as well as methods used for missing data: • n/a	
1. Were there a reasonable number of outcome events?	Yes
2. Was the time interval between predictor assessment and outcome determination appropriate?	Yes
3. Were all enrolled participants included in the analysis?	No
4. Were participants with missing data handled appropriately?	n/a
Risk of bias introduced by sample size or participant flow	RISK: High

DOMAIN 5: **Analysis**

A. Risk of Bias

Describe how the model was developed (predictor selection, fitting and optimism, risk groups, model performance): • General linear model (GLM) used to describe how change in function is affected by fear avoidance Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): • n/a Describe the performance measures of the model, e.g. calibration, discrimination, (re)classification, net benefit: • n/a	
1. Were non-binary predictors handled appropriately?	Yes
2. Was selection of predictors based on univariable analysis avoided?	Yes
3. Was model overfitting (optimism in model performance) accounted for, e.g. using bootstrapping or shrinkage techniques?	No
4. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	No
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	No
6. Were relevant model performance measures evaluated, e.g. calibration, discrimination, (re)classification and net benefit?	No
Risk of bias introduced by the analysis	RISK: High

Step 4: Overall judgement

Overall judgement about risk of bias and applicability of the prediction model evaluation

Overall judgement of risk of bias	RISK: High
Overall judgement of applicability	CONCERN: High

Step 5: Usability of the model

Assess the usability of the model	
Is the model presented with sufficient detail to be used in the intended context and target population?	RATING: No

Tyler et al. (2010)

Step 3: Assess risk of bias and applicability

DOMAIN 1: **Participant selection**

A. Risk of Bias

Describe the sources of data and criteria for participant selection: • Prospective cohort study • 'Internal impingement' Diagnostic criteria for internal impingement was used. Specific inclusion criteria based on physical examination were positive relocation test, positive posterior impingement sign, and posterior glenohumeral joint line tenderness. Specific inclusion criteria based on MRI findings were the presence of a posterosuperior glenoid labral lesion. Exclusion criteria were anterior instability, full-thickness rotator cuff tear, and subacromial impingement as determined by physical examination and MRI. Additionally, all patients reported subjective clicking in their shoulder on active movement.	
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Yes
2. Were all inclusions and exclusions of participants appropriate?	Yes
3. Were participants enrolled at a similar state of health, or were predictors considered to account for any dissimilarities?	Uncertain
Risk of bias introduced by selection of participants	RISK: Low

B. Applicability

Describe included participants and setting: n=22 No details	
Concern that the included participants and setting do not match the review question	CONCERN: High

DOMAIN 2: **Predictors**

A. Risk of Bias

List and describe predictors assessed, e.g. definition, timing and knowledge of other predictors: • Demographics factors: None • Clinical history factors: None • Patient reported measures: Pain (none), Psychological symptoms (none), Function / Disability (none) • Clinical measures: Strength (none), ROM (Improvement in PST at discharge, Improvement in passive ER ROM, Improvement in GIRD), Scapular movement and control (none) • Structural pathology: Imaging (none), Orthopaedic tests (none) All recorded at initiation of the study	
1. Were predictors defined and assessed in a similar way for all participants?	Yes
3. Were predictor assessments made without knowledge of outcome data?	Yes
4. Are all predictors available at the time the model is intended to be used?	No
5. Were all relevant predictors analysed?	Yes
Risk of bias introduced by predictors or their assessment	RISK: High

B. Applicability

Concern that the definition, assessment or timing of assessment of predictors in the model do not match the review question	CONCERN: High

DOMAIN 3: **Outcome**

A. Risk of Bias

Describe the outcome and how it was defined and determined: Symptom free via Simple Shoulder Test at discharge; upon completion of treatment	
1. Was a pre-specified outcome definition used?	Yes
2. Were predictors excluded from the outcome definition?	Yes
3. Was the outcome defined and determined in a similar way for all participants?	Yes
5. Was the outcome determined without knowledge of predictor information?	Yes
Risk of bias introduced by the outcome or its determination	RISK: Low

B. Applicability

At what time point was the outcome determined: Upon completion of treatment

If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome: n/a	
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: Low

DOMAIN 4: **Sample size and participant flow**

A. Risk of Bias

Describe numbers of participants and events per predictor: • n=22; 7 Describe the time interval between predictor assessment and outcome determination: • Upon completion of treatment Describe any participants who were excluded from the model: • No data Describe missing data on predictors and outcomes as well as methods used for missing data: • n/a	
1. Were there a reasonable number of outcome events?	No
2. Was the time interval between predictor assessment and outcome determination appropriate?	Yes
3. Were all enrolled participants included in the analysis?	NI
4. Were participants with missing data handled appropriately?	n/a
Risk of bias introduced by sample size or participant flow	RISK: High

DOMAIN 5: **Analysis**

A. Risk of Bias

Describe how the model was developed (predictor selection, fitting and optimism, risk groups, model performance): • Mixed model analysis of variance, with Treatment (pretreatment vs posttreatment) as the within-subjects factor and Group (patients with complete resolution of symptoms vs patients with residual symptoms) as the between-subjects factor Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): • n/a Describe the performance measures of the model, e.g. calibration, discrimination, (re)classification, net benefit: • n/a	
1. Were non-binary predictors handled appropriately?	Yes

2. Was selection of predictors based on univariable analysis avoided?	Yes
3. Was model overfitting (optimism in model performance) accounted for, e.g. using bootstrapping or shrinkage techniques?	No
4. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	No
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	NI
6. Were relevant model performance measures evaluated, e.g. calibration, discrimination, (re)classification and net benefit?	No
Risk of bias introduced by the analysis	RISK: High

Step 4: Overall judgement

Overall judgement about risk of bias and applicability of the prediction model evaluation	
Overall judgement of risk of bias	RISK: High
Overall judgement of applicability	CONCERN: High

Step 5: Usability of the model

Assess the usability of the model	
Is the model presented with sufficient detail to be used in the intended context and target population?	RATING: No

Virta et al. (2009)

Step 3: Assess risk of bias and applicability

DOMAIN 1: **Participant selection**

A. Risk of Bias

Describe the sources of data and criteria for participant selection: • Prospective cohort study • 'Subacromial impingement' Diagnosis was confirmed with subacromial anaesthesia, with a few exceptions by the same doctor, and most of the patients had passed an MRI examination. Patients were included if diagnosis was confirmed.	
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Yes
2. Were all inclusions and exclusions of participants appropriate?	Yes
3. Were participants enrolled at a similar state of health, or were predictors considered to account for any dissimilarities?	Uncertain
Risk of bias introduced by selection of participants	RISK: Low

B. Applicability

Describe included participants and setting: n=97 Recruited patients who were on a waiting list for orthopaedic surgery but had been referred for physiotherapy prior to surgery	
Concern that the included participants and setting do not match the review question	CONCERN: High

DOMAIN 2: **Predictors**

A. Risk of Bias

List and describe predictors assessed, e.g. definition, timing and knowledge of other predictors: • Demographics factors: Age, gender • Clinical history factors: Symptom duration • Patient reported measures: Pain (none), Psychological symptoms (none), Function / Disability (none) • Clinical measures: Strength (none), ROM (none), Scapular movement and control (none) • Structural pathology: Imaging (none), Orthopaedic tests (none) All recorded at initiation of the study	
1. Were predictors defined and assessed in a similar way for all participants?	Yes
3. Were predictor assessments made without knowledge of outcome data?	Yes

4. Are all predictors available at the time the model is intended to be used?	Yes
5. Were all relevant predictors analysed?	Yes
Risk of bias introduced by predictors or their assessment	RISK: Low

B. Applicability

Concern that the definition, assessment or timing of assessment of predictors in the model do not match the review question	CONCERN: Low

DOMAIN 3: **Outcome**

A. Risk of Bias

Describe the outcome and how it was defined and determined: UCLA; upon completion of treatment	
1. Was a pre-specified outcome definition used?	Yes
2. Were predictors excluded from the outcome definition?	Yes
3. Was the outcome defined and determined in a similar way for all participants?	Yes
5. Was the outcome determined without knowledge of predictor information?	Yes
Risk of bias introduced by the outcome or its determination	RISK: Low

B. Applicability

At what time point was the outcome determined: Upon completion of treatment If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome: n/a	
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: Low

DOMAIN 4: **Sample size and participant flow**

A. Risk of Bias

Describe numbers of participants and events per predictor: • n=97; 18 Describe the time interval between predictor assessment and outcome determination: • Upon completion of treatment Describe any participants who were excluded from the model: • 26% but failed to compare the baseline characteristics of those who were and those who were not followed up

Describe missing data on predictors and outcomes as well as methods used for missing data: • n/a	
1. Were there a reasonable number of outcome events?	Yes
2. Was the time interval between predictor assessment and outcome determination appropriate?	Yes
3. Were all enrolled participants included in the analysis?	Yes
4. Were participants with missing data handled appropriately?	n/a
Risk of bias introduced by sample size or participant flow	RISK: Low

DOMAIN 5: **Analysis**

A. Risk of Bias

Describe how the model was developed (predictor selection, fitting and optimism, risk groups, model performance): • No attempt made to trim the number of candidate variables • No details of analysis Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): • n/a Describe the performance measures of the model, e.g. calibration, discrimination, (re)classification, net benefit: • n/a	
1. Were non-binary predictors handled appropriately?	Yes
2. Was selection of predictors based on univariable analysis avoided?	NI
3. Was model overfitting (optimism in model performance) accounted for, e.g. using bootstrapping or shrinkage techniques?	No
4. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	No
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	NI
6. Were relevant model performance measures evaluated, e.g. calibration, discrimination, (re)classification and net benefit?	No
Risk of bias introduced by the analysis	RISK: High

Step 4: Overall judgement

Overall judgement about risk of bias and applicability of the prediction model evaluation	
Overall judgement of risk of bias	RISK: High
Overall judgement of applicability	CONCERN: High

Step 5: Usability of the model

Assess the usability of the model	
Is the model presented with sufficient detail to be used in the intended context and target population?	RATING: No

10.4 PROBAST Overall judgement

Study	Risk of bias					Applicability concerns			Overall judgements		
	Participant selection	Predictors	Outcome	Sample size and flow	Analysis	Participant selection	Predictors	Outcome	Risk of bias	Applicability	Usability of the model
Bartolozzi et al. (1994)	High	Low	High	High	High	Low	Low	Low	High	Low	No
Chester et al (2016)	Low	Low	High	Low	High	High	Low	Low	High	High	No
Conroy and Hayes (1998)	Low	Low	Low	High	High	High	Low	Low	High	High	No
Deutscher et al. (2009)	High	Low	High	High	High	High	Low	Low	High	High	No
Engebretsen et al. (2010)	Low	Low	High	High	High	Low	Low	Low	High	Low	No
Hung et al (2010)	Low	High	Low	High	High	Low	High	Low	High	High	No
Kennedy et al. (2006)	Low	Low	High	High	High	High	Low	Low	High	High	No
Kromer et al (2014)	Low	Low	High	High	High	Low	Low	Low	High	Low	No
Mintken et al. (2010)	Low	Low	Low	High	High	Low	Low	High	High	High	No
Morrison et al. (1997)	Low	High	Low	High	High	Low	Low	Low	High	Low	No
Ogon et al. (2009)	Low	High	Low	Uncertain	High	High	High	High	High	High	No

Sindhu et al. (2012)	Low	High	Low	High	High	High	High	Low	High	High	No
Tyler et al. (2010)	Low	High	Low	High	High	High	High	Low	High	High	No
Virta et al. (2009)	Low	Low	Low	Low	High	High	Low	Low	High	High	No

11 Appendix IV: PreMethods chapter

GRRAS checklist for reporting of studies of reliability and agreement

D1. Clinical measures: Strength
- Demographics of strength (and ROM) intra-rater reliability study subjects
- Strength (and ROM) intra-rater reliability study data collection form
- Strength intra-rater reliability study: Raw data

D2. Clinical measures: ROM
- ROM intra-rater reliability study: Raw data

D3. Clinical measures: Scapular movement and control
- Demographics of scapular movement and control inter-rater reliability study subjects
- Scapular dyskinesis gradings for each rater: raw data

E1. Structural pathology via imaging
- Diagnostic ultrasound PgC training
- Shoulder ultrasound reproducibility study patient information sheet
- Shoulder ultrasound reproducibility study patient consent sheet
- European Society of Skeletal Radiology technical guidelines for the shoulder
- Ultrasound differential diagnoses and record scan findings
- Demographics and clinical data of ultrasound inter-rater reliability study subjects
- Structural pathology findings from ultrasound inter-rater reliability study

Treatment
- Preliminary version of proforma for collecting patient treatment information
- Final version of proforma for collecting patient treatment information

11.1 GRRAS checklist for reporting of studies of reliability and agreement

Version based on Table I in: Kottner, J. et al. 2011. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J Clin Epidemiol. 64(1):96-106

4.1.1 D1. Clinical measures: Strength

Section	Item #	Checklist item	Details
Title / Abstract	1	Identify in title or abstract that interrater/intrarater reliability or agreement was investigated.	Strength intra-rater reliability study
Introduction	2	Name and describe the diagnostic or measurement device of interest explicitly.	Handheld dynamometer (Commander Muscle Tester, JTech Medical, USA)
	3	Specify the subject population of interest.	12 healthy subjects
	4	Specify the rater population of interest (if applicable).	A single rater (MS) who performed the prognostic cohort study strength measurements
	5	Describe what is already known about reliability and agreement and provide a rationale for the study (if applicable).	Comparable study by Awatani et al. (2016); ICC between 0.850 and 0.980 As a single rater (MS) performed the prognostic cohort study strength measurements, it was deemed necessary to assess the absolute accuracy of the strength measures taken in the same test positions using the same measurement tool as for the prognostic cohort study
Methods	6	Explain how the sample size was chosen. State the determined number of raters, subjects/objects, and replicate observations.	Sample size chosen to align with previous comparable publication (Awatani et al. 2016). Raters = 1 Subjects = 12 Replicate observations = 2
	7	Describe the sampling method.	Convenience sample
	8	Describe the measurement/rating process (e.g. time interval between repeated measurements, availability of clinical information, blinding).	Time interval between repeated measurements = within session; precise timing not controlled Availability of clinical information = n/a Blinding = data stored on dynamometer; rater did not observe the readings during testing.
	9	State whether measurements/ratings were conducted independently.	Measurements were taken independently of the previous value.
	10	Describe the statistical analysis.	Intraclass correlation coefficient ($ICC_{3,1}$)
Results	11	State the actual number of raters and subjects/objects which were included and the	Raters = 1 Subjects = 12 Replicate observations = 2

		number of replicate observations which were conducted.	
	12	Describe the sample characteristics of raters and subjects (e.g. training, experience).	Rater = experienced MSK clinician and researcher Subjects = n/a
	13	Report estimates of reliability and agreement including measures of statistical uncertainty.	See table 4.3 for ICC(3,1) results, including 95 % CI
Discussion	14	Discuss the practical relevance of results.	The intra-rater reliability findings for the left and right side for IR, ER and Scaption are all in the very good range (Landis and Koch 1977; Altman 1991). Therefore the data collection method can be considered to be reliable and so the tool could be used in the same format in the prognostic cohort study.
Auxiliary material	15	Provide detailed results if possible (e.g. online).	See appendix 11 , tables 11.2 to 11.7

4.1.2 D2. Clinical measures: ROM

Section	Item #	Checklist item	Details
Title / Abstract	1	Identify in title or abstract that interrater/intrarater reliability or agreement was investigated.	ROM intra-rater reliability study
Introduction	2	Name and describe the diagnostic or measurement device of interest explicitly.	Digital inclinometer (DigiPas DWL-180s, JSB Tech Pte Ltd, Singapore)
	3	Specify the subject population of interest.	12 healthy subjects
	4	Specify the rater population of interest (if applicable).	A single rater (MS) who performed the prognostic cohort study ROM measurements
	5	Describe what is already known about reliability and agreement and provide a rationale for the study (if applicable).	Comparable study by Mullaney et al. (2010); ICC between 0.94 and 0.98 As a single rater (MS) performed the prognostic cohort study ROM measurements, it was deemed necessary to assess the absolute accuracy of the ROM measures taken in the same test positions using the same measurement tool as for the prognostic cohort study
Methods	6	Explain how the sample size was chosen. State the determined number of raters, subjects/objects, and replicate observations.	Sample size chosen to align with previous comparable publication (Mullaney et al. 2010) and make efficient use of strength intra-rater reliability study sample. Raters = 1 Subjects = 12 Replicate observations = 2
	7	Describe the sampling method.	Convenience sample
	8	Describe the measurement/rating process (e.g. time interval between repeated measurements, availability of clinical information, blinding).	Time interval between repeated measurements = within session; precise timing not controlled Availability of clinical information = n/a Blinding = none.
	9	State whether measurements/ratings were conducted independently.	Measurements were taken independently of the previous value.
	10	Describe the statistical analysis.	Intraclass correlation coefficient ($ICC_{3,1}$)
Results	11	State the actual number of raters and subjects/objects which were included and the number of replicate observations which were conducted.	Raters = 1 Subjects = 12 Replicate observations = 2
	12	Describe the sample characteristics of raters and	Rater = experienced MSK clinician and researcher

		subjects (e.g. training, experience).	Subjects = n/a
	13	Report estimates of reliability and agreement including measures of statistical uncertainty.	See table 4.5 for ICC(3,1) results, including 95 % CI
Discussion	14	Discuss the practical relevance of results.	The intra-rater reliability findings for both <90° elevation and >90° elevation are both in the very good range (Landis and Koch 1977; Altman 1991). Therefore the data collection method can be considered to be reliable and so the tool could be used in the same format in the prognostic cohort study.
Auxiliary material	15	Provide detailed results if possible (e.g. online).	See appendix 11 , table 11.8

4.1.3 D3. Clinical measures: scapular movement and control

Section	Item #	Checklist item	Details
Title / Abstract	1	Identify in title or abstract that interrater/intrarater reliability or agreement was investigated.	Inter-rater reliability of the Scapular Dyskinesis Test (SDT)
Introduction	2	Name and describe the diagnostic or measurement device of interest explicitly.	The scapular dyskinesis test (SDT) which is a qualitative assessment of scapular position and movement based upon descriptive criteria and subsequently rating patients according to normal, subtle or obvious dyskinesis (McClure et al. 2009).
	3	Specify the subject population of interest.	The reliability study was undertaken on a subset of patients (n=30) from the main cohort
	4	Specify the rater population of interest (if applicable).	Level of agreement in rating the symptomatic side was established with one of the treating musculoskeletal clinicians in the study. This was Kevin Nicholas (KN) who worked at the Whitchurch hospital (WHI) site and had 10 years' experience working in musculoskeletal outpatients.
	5	Describe what is already known about reliability and agreement and provide a rationale for the study (if applicable).	Comparable study by McClure et al. (2009); Right side Kappa = 0.61, left side Kappa = 0.48; both weighted Kappa The SDT was developed on overhead athletes whilst the current study is concerned with NHS patients who were a non-athletic population. The likely greater heterogeneity of NHS patients in terms of body habitus and range of movement at the shoulder complex meant that establishing the level of agreement between the MS and another experienced musculoskeletal clinician using a sample of such patients was necessary
Methods	6	Explain how the sample size was chosen. State the determined number of raters, subjects/objects, and replicate observations.	Sample size chosen to align with previous comparable study by McClure et al. (2009); where each pair of raters rated 30 different participants. Raters = 2 Subjects = 30 Replicate observations = 2
	7	Describe the sampling method.	30 subjects who were not used in the training phase and who were not treated at the same site where KN worked.

	8	Describe the measurement/rating process (e.g. time interval between repeated measurements, availability of clinical information, blinding).	Time interval between repeated measurements = n/a; measurements performed in different locations Availability of clinical information = KN blinded to clinical information Blinding = absolute; measurements performed in different locations.
	9	State whether measurements/ratings were conducted independently.	Measurements were taken independently as measurements performed in different locations.
	10	Describe the statistical analysis.	Weighted Kappa was also performed; 0, 1, 2 for same grade, one grade disparity and two grades disparity; respectively.
Results	11	State the actual number of raters and subjects/objects which were included and the number of replicate observations which were conducted.	Raters = 2 Subjects = 30 Replicate observations = 2
	12	Describe the sample characteristics of raters and subjects (e.g. training, experience).	Rater = KN and MS undertook standardised training, which comprised the operational definitions, rating scale and photographic examples provided in the original publication. As the SDT is based upon interpretation of descriptive terminology, it was considered essential that the training and familiarisation process included applying the SDT to footage of scapular movement. Twenty videos of scapular movement patterns from patients with SIS/RCTendinopathy from both loaded (i.e. holding a weight in each hand) and un-loaded (i.e. not holding a weight) trials which demonstrated a spectrum of scapular movement patterns were therefore selected from the prognostic cohort. KN and MS viewed these videos and had the opportunity to discuss the interpretation and application of the descriptive terminology and recording process. This mirrored a typical, peer-supported clinical training environment (Baertschi et al. 2013). Subjects = n/a
	13	Report estimates of reliability and agreement including measures of statistical uncertainty.	Kappa = 0.329, $p<0.002$
Discussion	14	Discuss the practical relevance of results.	The linear weighted Kappa results confirm fair agreement (Kappa 0.21 to 0.40) (Viera

			and Garrett 2005) for inter-rater reliability of Scapular grading. Therefore the SDT must be applied with caution in the main cohort study.
Auxiliary material	15	Provide detailed results if possible (e.g. online).	See appendix 11 , table 11.10

4.2.1 E1. Structural pathology via imaging

Section	Item #	Checklist item	Details
Title / Abstract	1	Identify in title or abstract that interrater/intrarater reliability or agreement was investigated.	Shoulder ultrasound reproducibility study
Introduction	2	Name and describe the diagnostic or measurement device of interest explicitly.	Diagnostic ultrasound scan performed by Sonographer
	3	Specify the subject population of interest.	Sample of patients referred for ultrasound scans of their shoulder in the Radiology department at the University Hospital of Wales
	4	Specify the rater population of interest (if applicable).	Reproducibility study was undertaken comparing scans performed by MS against an experienced musculoskeletal ultrasound practitioner, Dr Peter Mullaney (PM). PM has been a Consultant Radiologist since 2007 and is a level 3 ultrasound practitioner (The Royal College of Radiologists 2012).
	5	Describe what is already known about reliability and agreement and provide a rationale for the study (if applicable).	Comparable study by Thoomes-de Graaf et al. (2014); 65 patients were scanned by 1 physical therapist and 1 radiologist (total of 13 physical therapists and 9 radiologists). As a highly operator dependent modality it was deemed necessary for MS to undergo formal training in diagnostic ultrasound and to undertake an inter-rater reliability study.
Methods	6	Explain how the sample size was chosen. State the determined number of raters, subjects/objects, and replicate observations.	Sample size chosen to align with previous comparable study by Thoomes-de Graaf et al. (2014); sample size also influenced by pragmatic limitations, i.e. the option to include a particular subject in the study was dependent upon clinical service demands during that particular clinical session. Raters = 2 Subjects = 35 Replicate observations = 2
	7	Describe the sampling method.	Convenience sample
	8	Describe the measurement/rating process (e.g. time interval between	Time interval between repeated measurements = within session; precise timing not controlled Availability of clinical information = neither rater was blinded to clinical information

		repeated measurements, availability of clinical information, blinding).	Blinding = MS blinded to PM diagnosis
	9	State whether measurements/ratings were conducted independently.	MS diagnosis performed independently of PM.
	10	Describe the statistical analysis.	Kappa
Results	11	State the actual number of raters and subjects/objects which were included and the number of replicate observations which were conducted.	Raters = 2 Subjects = 35 Replicate observations = 2
	12	Describe the sample characteristics of raters and subjects (e.g. training, experience).	MS: School of Medicine in Cardiff University 2 year Postgraduate Certificate (PgC) in Medical ultrasound was undertaken as part of the training component of his PhD. The training was extensive and included formal assessments of theoretical knowledge, image optimisation and sonographic diagnosis (http://www.case-uk.org/; accessed 24/04/2017). Further details of the course requirements are provided in appendix 4. MS successfully completed the PgC (Distinction grade) and is on the voluntary register of Sonographers operated by the Society of Radiography (http://www.sor.org/practice/ultrasound/register-sonographers ; accessed 24/04/2017). PM: an experienced musculoskeletal ultrasound practitioner, Dr Peter Mullaney (PM). PM has been a Consultant Radiologist since 2007 and is a level 3 ultrasound practitioner (The Royal College of Radiologists 2012). Subjects = n/a
	13	Report estimates of reliability and agreement including measures of statistical uncertainty.	See table 4.13
Discussion	14	Discuss the practical relevance of results.	Based upon the Kappa values it can therefore be concluded that when compared to an experienced musculoskeletal radiologist, MS demonstrated acceptable agreement for all categories except for the sonographic diagnosis of tendinopathy.
Auxiliary material	15	Provide detailed results if possible (e.g. online).	See appendix 11 , table 11.12

11.2 Demographics of strength (and ROM) intra-rater reliability study subjects

Table 11-1: Demographics of strength (and ROM) intra-rater reliability study subjects

Subject number	Gender 1=M, 2=F	Age (years)	Height (metres)	Weight (kg)	BMI (kg/m^2)	Arm Dom 1=L, 2=R	OSS
1	2	40	1.71	69.6	23.8	2	48
2	2	67	1.72	65.3	22.1	2	46
3	1	38	1.79	83.2	26.0	2	48
4	1	27	1.91	93.6	25.7	2	48
5	2	55	1.57	63.3	25.7	2	45
6	2	37	1.67	67.9	24.3	2	48
7	1	40	1.89	78.8	22.1	2	48
8	1	39	1.83	93.4	27.9	2	48
9	2	27	1.60	81.3	31.8	2	48
10	2	43	1.69	71.6	25.1	1	48
11	2	29	1.59	66.7	26.4	2	48
12	2	21	1.73	65.5	21.9	2	47

Key: M = male, F = female, Kg = kilograms, m = metres, BMI = body mass index, L = left, R = right, OSS = Oxford shoulder score

11.3 Strength (and ROM) intra-rater reliability study data collection form

Subject (Initial + date) Gender (1=M, 2=F) Age

Height Weight Arm dominance (1=L, 2=R, 3=ambidextrous)

Inclinometer

< 90 = L or R > 90 = L or R

	Left	Right
Trial 1		
Trial 2		
Trial 3		

HHD

Distance elbow to radial styloid = Distance GHJ to radial styloid =

IR

	Left	Right
Trial 1		
Trial 2		
Trial 3		

ER

	Left	Right
Trial 1		
Trial 2		
Trial 3		

Scaption

	Left	Right
Trial 1		
Trial 2		
Trial 3		

11.4 Strength intra-rater reliability study: Raw data

Table 11-2: Strength intra-rater reliability study: Raw data – Left internal rotation

Subject number	Length of lever arm (m)	Left internal rotation Force (N)			Left internal rotation Moment (Nm)		
		Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
1	0.26	79.2	74.8	70.4	20.6	19.4	18.3
2	0.27	44.0	44.0	44.0	11.9	11.9	11.9
3	0.25	140.0	140.0	140.0	35.0	35.0	35.0
4	0.27	99.0	101.0	101.0	26.7	27.3	27.3
5	0.22	44.0	55.0	44.0	9.7	12.1	9.7
6	0.24	99.0	101.0	105.0	23.8	24.2	25.2
7	0.29	105.0	114.0	105.0	30.5	33.1	30.5
8	0.28	105.0	94.6	105.0	29.4	26.5	29.4
9	0.24	88.0	79.2	79.2	21.1	19.0	19.0
10	0.23	61.6	59.4	61.6	14.2	13.7	14.2
11	0.25	70.4	63.8	79.2	17.6	16.0	19.8
12	0.27	61.6	55.0	66.0	16.6	14.9	17.8

Key: m = metres, N = Newtons, Nm = Newton metres

Table 11-3: Strength intra-rater reliability study: Raw data – Right internal rotation

Subject number	Length of lever arm (m)	Right internal rotation Force (N)			Right internal rotation Moment (Nm)		
		Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
1	0.26	79.2	79.2	77.0	20.6	20.6	20.0
2	0.27	41.8	44.0	44.0	11.3	11.9	11.9
3	0.25	149.0	140.0	132.0	37.3	35.0	33.0
4	0.27	105.0	99.0	90.2	28.4	26.7	24.4
5	0.22	39.6	44.0	44.0	8.7	9.7	9.7
6	0.24	105.0	96.8	90.2	25.2	23.2	21.6
7	0.29	101.0	105.0	105.0	29.3	30.5	30.5
8	0.28	101.0	114.0	123.0	28.3	31.9	34.4
9	0.24	79.2	77.0	77.0	19.0	18.5	18.5
10	0.23	44.0	44.0	44.0	10.1	10.1	10.1
11	0.25	61.6	70.4	70.4	15.4	17.6	17.6
12	0.27	59.4	59.4	57.2	16.0	16.0	15.4

Key: m = metres, N = Newtons, Nm = Newton metres

Table 11-4: Strength intra-rater reliability study: Raw data - Left external rotation

Subject number	Length of lever arm (m)	Left external rotation Force (N)			Left external rotation Moment (Nm)		
		Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
1	0.26	70.4	79.2	79.2	18.3	20.6	20.6
2	0.27	59.4	44.0	59.4	16.0	11.9	16.0
3	0.25	70.4	70.4	70.4	17.6	17.6	17.6
4	0.27	94.6	74.8	105.0	25.5	20.2	28.4
5	0.22	44.0	44.0	44.0	9.7	9.7	9.7
6	0.24	44.0	55.0	44.0	10.6	13.2	10.6
7	0.29	79.2	79.2	70.4	23.0	23.0	20.4
8	0.28	96.8	101.0	114.0	27.1	28.3	31.9
9	0.24	61.6	61.6	66.0	14.8	14.8	15.8
10	0.23	52.8	52.8	46.2	12.1	12.1	10.6
11	0.25	70.4	79.2	66.0	17.6	19.8	16.5
12	0.27	50.6	66.0	52.8	13.7	17.8	14.3

Key: m = metres, N = Newtons, Nm = Newton metres

Table 11-5: Strength intra-rater reliability study: Raw data - Right external rotation

Subject number	Length of lever arm (m)	Right external rotation Force (N)			Right external rotation Moment (Nm)		
		Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
1	0.26	74.8	70.4	70.4	19.4	18.3	18.3
2	0.27	61.6	52.8	46.2	16.6	14.3	12.5
3	0.25	61.6	66.0	70.4	15.4	16.5	17.6
4	0.27	96.8	99.0	99.0	26.1	26.7	26.7
5	0.22	44.0	44.0	44.0	9.7	9.7	9.7
6	0.24	55.0	61.6	63.8	13.2	14.8	15.3
7	0.29	99.0	88.0	101.0	28.7	25.5	29.3
8	0.28	105.0	105.0	105.0	29.4	29.4	29.4
9	0.24	63.8	61.6	59.4	15.3	14.8	14.3
10	0.23	61.6	48.4	61.6	14.2	11.1	14.2
11	0.25	66.0	66.0	70.4	16.5	16.5	17.6
12	0.27	52.8	52.8	44.0	14.3	14.3	11.9

Key: m = metres, N = Newtons, Nm = Newton metres

Table 11-6: Strength intra-rater reliability study: Raw data - Left scaption

Subject number	Length of lever arm (m)	Left scaption Force (N)			Left scaption Moment (Nm)		
		Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
1	0.50	55.0	52.8	61.6	27.5	26.4	30.8
2	0.60	28.6	19.8	26.4	17.2	11.9	15.8
3	0.52	59.4	66.0	70.4	30.9	34.3	36.6
4	0.54	57.2	44.0	61.6	30.9	23.8	33.3
5	0.44	37.4	44.0	44.0	16.5	19.4	19.4
6	0.50	55.0	46.2	52.8	27.5	23.1	26.4
7	0.52	44.0	44.0	61.6	22.9	22.9	32.0
8	0.53	66.0	66.0	61.6	35.0	35.0	32.6
9	0.47	70.4	70.4	66.0	33.1	33.1	31.0
10	0.47	35.2	35.2	35.2	16.5	16.5	16.5
11	0.48	44.0	44.0	48.4	21.1	21.1	23.2
12	0.52	35.2	35.2	35.2	18.3	18.3	18.3

Key: m = metres, N = Newtons, Nm = Newton metres

Table 11-7: Strength intra-rater reliability study: Raw data - Right scaption

Subject number	Length of lever arm (m)	Right scaption Force (N)			Right scaption Moment (Nm)		
		Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
1	0.50	44.0	46.2	61.6	22.0	23.1	30.8
2	0.60	19.8	19.8	24.2	11.9	11.9	14.5
3	0.52	55.0	70.4	70.4	28.6	36.6	36.6
4	0.54	63.8	61.6	61.6	34.5	33.3	33.3
5	0.44	35.2	41.8	44.0	15.5	18.4	19.4
6	0.50	63.8	63.8	70.4	31.9	31.9	35.2
7	0.52	30.8	35.2	44.0	16.0	18.3	22.9
8	0.53	70.4	61.6	61.6	37.3	32.6	32.6
9	0.47	61.6	61.6	66.0	29.0	29.0	31.0
10	0.47	30.8	33.0	35.2	14.5	15.5	16.5
11	0.48	44.0	55.0	57.2	21.1	26.4	27.5
12	0.52	44.0	35.2	44.0	22.9	18.3	22.9

Key: m = metres, N = Newtons, Nm = Newton metres

11.5 ROM intra-rater reliability study: Raw data

Table 11-8: ROM intra-rater reliability study: Raw data

Subject number	< 90° (°)			> 90° (°)		
	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
1	50.0	47.4	49.4	144.2	144.6	139.8
2	41.6	39.4	44.0	128.3	128.3	130.5
3	39.4	40.3	37.2	139.8	136.5	133.6
4	33.2	31.8	32.8	124.5	123.2	119.5
5	62.3	62.1	61.5	131.1	136.6	136.9
6	53.4	51.8	52.6	135.2	133.4	131.2
7	30.0	30.8	31.3	131.5	130.6	128.7
8	28.0	29.1	26.6	136.3	135.5	134.2
9	47.0	48.0	49.0	147.0	149.0	151.0
10	44.5	47.4	46.3	140.1	142.0	142.1
11	45.4	50.0	49.8	140.3	141.9	142.7
12	46.3	46.9	47.7	132.7	130.5	132.0

Key: ° = degrees

11.6 Demographics of scapular movement and control inter-rater reliability study subjects

Table 11-9: Demographics of scapular movement and control inter-rater reliability study subjects

Subject code	Gender 1=M, 2=F	Symptomatic side 1=L, 2=R	Weighted trial? 1=Y, 2=N	BMI /kgm^{-2}	Maximum range of movement on symptomatic side /°
F6847h	2	2	2	35.4	132
F7047h	2	2	2	26.3	150
M2146h	1	2	1	37.3	146
F5446h	2	2	1	18.2	163
F4417h	2	1	1	50.5	75
M0379h	1	1	2	26.4	145
F7849h	2	2	1	23.8	168
F4583h	2	2	2	26.5	94
M0825h	1	2	1	29.5	150
M0061h	1	2	2	31.2	146
F8486h	2	2	1	32.2	155
F2520h	2	1	2	40.3	115
M1586h	1	2	1	27.2	166
F0165h	2	2	2	25.6	148
M8878h	1	1	1	23.0	136
F6416h	2	2	2	27.4	137
F7405h	2	1	1	20.5	154
M9819h	1	2	2	23.0	134
F9939h	2	2	1	24.3	121
F0809h	2	1	2	32.0	107
M5703h	1	1	1	30.5	146
M1518h	1	1	2	31.0	138
M1243h	1	1	1	31.3	108
M3288h	1	1	2	31.3	142
F4943h	2	2	1	24.1	149
M0519h	1	1	2	26.7	150
F1669h	2	2	1	30.0	164
F9261h	2	1	2	25.6	74
F5690h	2	2	1	24.4	160
M9287h	1	2	2	33.1	174

Key: M = male, F = female, L = left, R = right, Y = yes, N = no, Kg = kilograms, m = metres, BMI = body mass index, ° = degrees

11.7 Scapular dyskinesis gradings for each rater: raw data

Table 11-10: Scapular dyskinesis gradings for each rater: raw data

Subject code	MS rating	KN rating
F6847h	2	2
F7047h	3	3
M2146h	3	3
F5446h	3	2
F4417h	3	3
M0379h	2	3
F7849h	2	2
F4583h	1	3
M0825h	2	2
M0061h	2	3
F8486h	2	2
F2520h	3	3
M1586h	2	2
F0165h	1	3
M8878h	2	2
F6416h	1	3
F7405h	3	1
M9819h	2	3
F9939h	1	3
F0809h	1	1
M5703h	3	3
M1518h	2	1
M1243h	2	3
M3288h	2	2
F4943h	1	1
M0519h	3	3
F1669h	2	2
F9261h	3	3
F5690h	1	1
M9287h	1	2

Key: 1 = Normal, 2 = Subtle abnormality, 3 = Obvious abnormality

11.8 Diagnostic ultrasound PgC training

The PgC was a CASE (Consortium for the Accreditation of Sonographic Education) approved course and included a foundation science module which covered the physics of ultrasound as applied to medicine, including image optimisation and safety implications. Assessment was through 4 essays (4 x 10% of module mark), 2 online assessments (2 x 10% of module mark) and a written examination (40% of module mark) was sat at the end of the first year.

The musculoskeletal ultrasound module covered the fundamentals of ultrasound examination of normal and pathological musculoskeletal tissue. It was assessed through a written examination (40% of module mark) and practical skills assessments covering communication, clinical evaluation, machine and imaging-related skills (40% of module mark). A log book of standardised images of the 2 anatomical regions (1 upper limb and 1 lower limb) which the researcher specialised in was also required: these were the shoulder and the foot and ankle, along with soft tissue differentiation (e.g. tumours, masses, swelling).

11.9 Shoulder ultrasound reproducibility study patient information sheet

PATIENT INFORMATION SHEET

Diagnostic ultrasound scanning of the shoulder

Version 2 30/01/2013

Part one

You are being invited to take part in a research study with Cardiff University's Arthritis Research Campaign Biomechanics and Bioengineering Centre. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. One of our team will go through the information sheet with you. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to participate. Part 1 tells you about the purpose of this study and what will happen to you if you take part. Part 2 gives you more detailed information about the conduct of the study.

What is the purpose of this trial?
The aim of the trial is to use an ultrasound scanner to look at the soft tissues (e.g. muscle, tendon, etc) of your shoulder. These tissues can be injured or damaged as part of the shoulder problem that you have. Such information is one of the factors that are being looked at to see if we can predict which people with shoulder pain can be helped with rehabilitation and which people are not helped with rehabilitation. Eventually this will help us to direct patients for the right treatment at the right time.

Do I have to take part?
It is up to you to whether or not to take part. If you do decide to take part you will be given this information sheet to keep and after you have had enough time to read through it, be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time or without giving a reason. A decision not to take part or to withdraw at any time will not affect the standard of care you receive. Should you decide not to take part, you do not have to provide a reason for this decision.

What will happen to me if I take part?
You have been asked to take part in this as you are a person with shoulder pain who have been referred to radiology or to physiotherapy.

If you wish to take part you will assessed on site in the NHS. If you have been referred for Physiotherapy then you may be invited to come back for subsequent scans at the Cardiff University School of Healthcare Studies (SOHCS) Research Centre for Clinical Kinaesiology (RCCK). The number of times we would ask you to attend will be discussed with you when going through this information sheet. The sessions will last a maximum of thirty minutes.

Data will be kept securely for a minimum of 15 years in accordance with good research practice and data protection regulations imposed by Cardiff University in accordance with the Data Protection Act 1998. All data obtained during the study will remain confidential. Access to data will only be available to the investigators attached to the Arthritis Research UK Biomechanics and Bioengineering Centre at Cardiff University.

If new information becomes available, we may invite you to take part in a follow-up study in the future, please indicate on the consent sheet if you do not mind us contacting you.

What will I have to do?
At the beginning of your visit, we will explain the full study to you and ask for your consent, bearing in mind that you are free to withdraw at any time.

Before your first assessment you will be asked to sign a consent form which includes the following clause: I understand that I may withdraw from the study at any time without it affecting any ongoing treatment in any way.

As with routine radiology and physiotherapy assessment of the shoulder, the shoulder and surrounding area need to be exposed so that the assessment can be accurately performed. You will be offered a special apron to wear that leaves your shoulder visible, but covers your chest.

For the scan procedure you will be seated with your hands either resting on your thighs in a comfortable position or your hands on your hips. A water-based contact medium (gel) will be applied to your skin in the shoulder region to allow for optimal transmission of ultrasound waves. An ultrasound transducer head will be placed on your skin and from this an image of the tissues of your shoulder will be generated and displayed on a screen. The position and movement of the transducer head, along with an audio-commentary of the scan, will be recorded using a digital camcorder. Following the assessment the water-based contact gel will be wiped off and you will be asked to return to your original clothing.

All data files, including audiovisual files will be stored in encrypted folders on Cardiff University password protected computers. Cardiff University and NHS members of staff who are directly involved with the study will have access to the files. The audiovisual files will be electronically destroyed up to 15 years from the commencement of the study.

Are there any risks in participating in this trial?
Ultrasound scanning is routinely used in clinical practice and as noted by the Safety Group of the British Medical Ultrasound Society (BMUS) “there is no evidence that diagnostic ultrasound has produced any harm to patients in the four decades that it has been in use” (http://www.bmus.org/policies-guides/pg-safety04.asp). Nonetheless, safety principles issued by BMUS will be strictly adhered to.

Are there any benefits in participating in this trial?
Whilst ultrasound can provide useful information about injured or damaged tissues around the shoulder, such findings can also occur in people who do not have any shoulder problems. If you are a patient who has been referred for Physiotherapy then your Doctor has referred you for rehabilitation based on a number of factors

However the ultrasound findings from this study will be one of multiple factors that will be looked at to see if we can identify and predict those patients with shoulder pain that can be helped with rehabilitation and which people are not helped with rehabilitation. Eventually this will help us to direct patients for the right treatment at the right time. If we do find an undiagnosed tear we will write to your GP and other clinicians and inform them.

If the information in Part 1 has interested you and you are considering participation, please read the additional information in Part 2 before making a decision.

Part Two

What if new information becomes available?
Sometimes during the course of a research project, new information becomes available about the investigation. If you decide to withdraw, it will not affect your any care in the NHS. If you decide to continue, you will be asked to sign an updated consent form.

What will happen if I do not want to carry on with the study?
If you withdraw from the study, we will erase all identifiable material, but we will need to use the data collected up to your withdrawal.

What if something goes wrong?
If you are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed due to someone's negligence, then you may have grounds for a legal action but you may have to pay for it. Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you have been approached or treated during the course of this study, the normal National Health Service complaints mechanisms should be available to you.

Will my taking part in this study be kept confidential?
All information which is collected about you during the course of the research will be kept strictly confidential. Any information about you which leaves the Cardiff University or the University Hospital of Wales will have your name and address removed so that you cannot be recognised from it. The video footage of the transducer position will be edited to ensure that you cannot be identified.

Will my GP be informed of my involvement in the study?
With your permission, we will send a letter to your General Practitioner informing him or her of your involvement in the study.

What will happen to the results of the research study?
The measurements taken will provide information about the movement of your joint. The results of the study will be presented at meetings of orthopaedic surgeons, clinical scientists, physiotherapists and engineers, and if accepted, published in medical and engineering journals. If interested, a copy of the published article can be made available to you. The video footage may be used in educational, clinical and research material and presentations. You will not be identified in any report/publication and the video footage will be anonymised so that you cannot be identified.

Who is organising and funding the research?
Research staff at the Arthritis Research UK Biomechanics and Bioengineering Centre at Cardiff University and Consultant Orthopaedic Surgeons at the University Hospital of Wales are carrying out the study. The study is part of the Arthritis Research UK Biomechanics and Bioengineering Centre at Cardiff University; it is not funded by commercial sources and runs alongside research in the Cardiff Gait and Motion Analysis Laboratory at Cardiff University School of Engineering and Research Centre for Clinical Kinaesiology at Cardiff University School of Healthcare Studies.

Who has reviewed the study?
This study has been reviewed by the Research Ethics Committee (REC) for Wales.

What if I wish to lodge a complaint?
If you wish to make a minor complaint regarding the way you were approached or treated during the trial, please contact the Arthritis Research UK Biomechanics and Bioengineering Centre Research

Coordinator at the contact details below or you can contact the Cardiff University Research Governance Team on 029 208 79277.

Contact for further information

Research Coordinator
Arthritis Research UK Biomechanics and Bioengineering Centre
Cardiff School of Biosciences
Cardiff University
Cardiff
CF10 3AX
Tel: 029 2087 5419
Email: Robertshc@cf.ac.uk or Longmanaj@cf.ac.uk

This completes Part 2. Thank you for reading this information sheet.

If you agree to take part in this study then you will be given a copy of the information sheet and a signed consent form to keep.

11.10 Shoulder ultrasound reproducibility study patient consent sheet

PATIENT CONSENT FORM

Diagnostic ultrasound scanning of the shoulder

Study Number
Patient Identification Number for this trial:

You DO NOT have to sign this document. Please DO NOT sign this document unless you fully understand it. If there is ANYTHING which you do not understand please do not hesitate to ask for a full explanation.

To confirm agreement with each of the statements below, please initial each box and amend as applicable:

1. I confirm that I have read and understand the information sheet dated 30/01/2013 (Version 2) for the above study and have had the opportunity to ask questions. ☐

2. I understand that my participation in the study is voluntary and that I am free to withdraw at any time, without giving any reason, and without my medical care or legal rights being affected. ☐

3. You may/may not contact me in the future to take part in other research projects or surveys ☐

4. I do/do not agree to my hospital number being used to track my data on your secure system. ☐

5. I do/do not agree to my GP and other clinicians being informed of my participation in the study. ☐

6. I agree to my scan being videoed using a digital camcorder ☐

7. I do/do not agree to the anonymised video images being used for clinical, educational and research purposes ☐

8. I agree to take part in the above study. ☐

Name of Patient: _____
(Please print)

Signature: _____ Date: _____
I confirm that I have fully explained the experimental protocol and purpose of the study

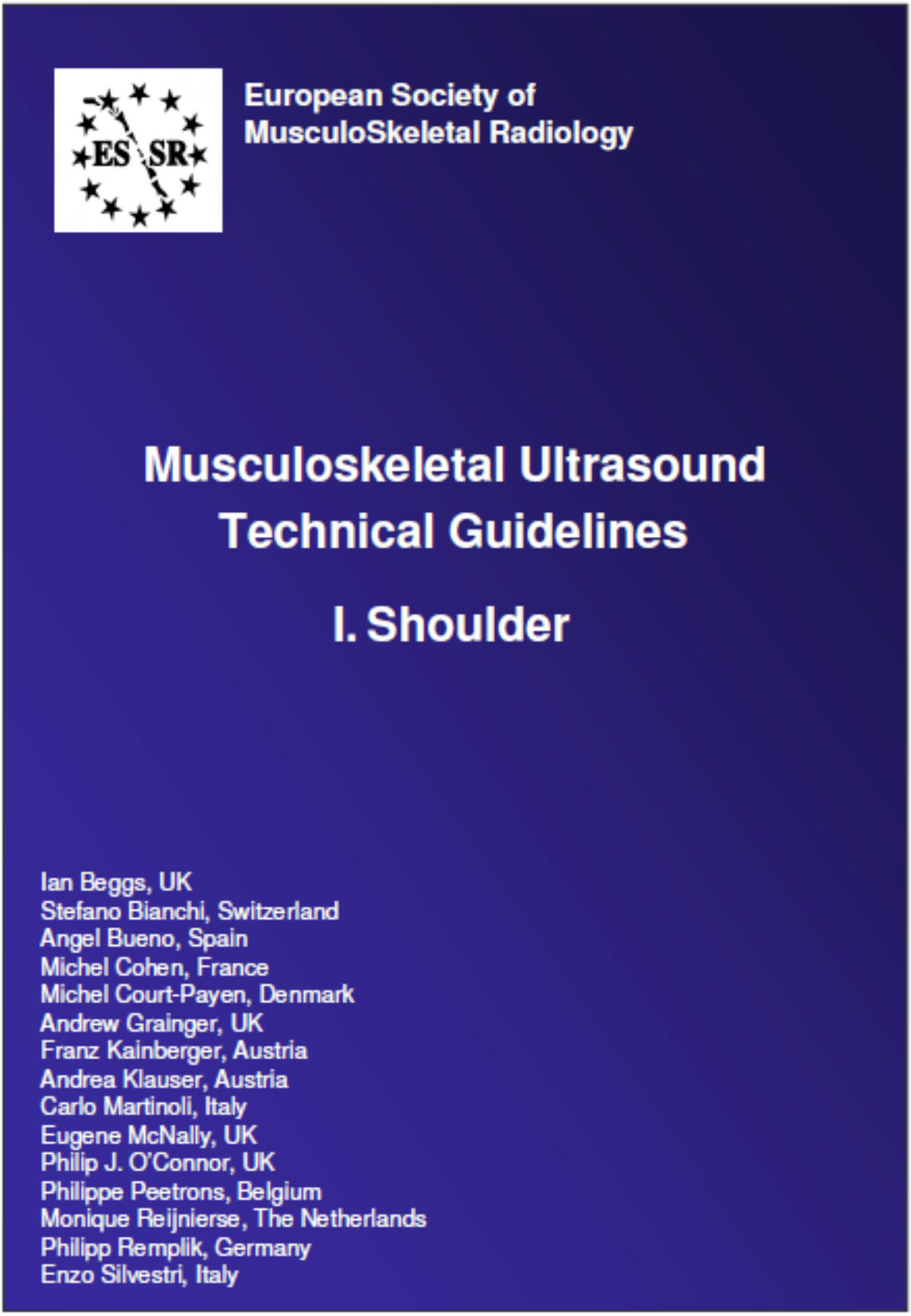
Name of Researcher: _____

Signature: _____ Date: _____

Name of person taking consent: _____
(If different from researcher)

Signature: _____ Date: _____

11.11 European Society of Skeletal Radiology technical guidelines for the shoulder

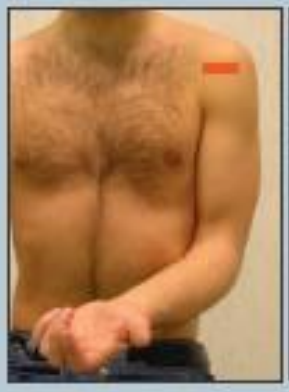


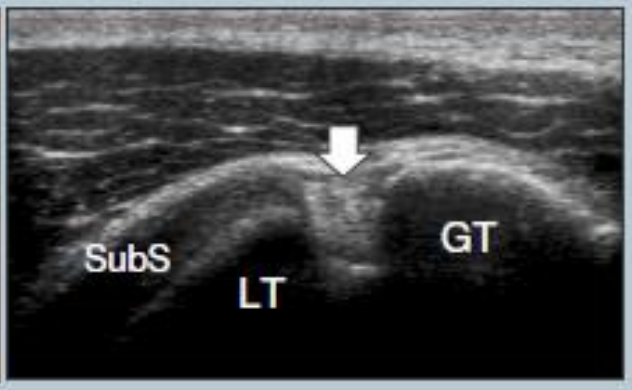
1

Although patient's positioning for shoulder US varies widely across different Countries and Institutions reflecting multifaceted opinions and experiences of different examiners, we strongly recommend to examine the patient while seated on a revolving stool. This position allows the examiner to reach the anterior, lateral and posterior aspects of the shoulder with the probe by simply asking the patient to rotate on the chair.

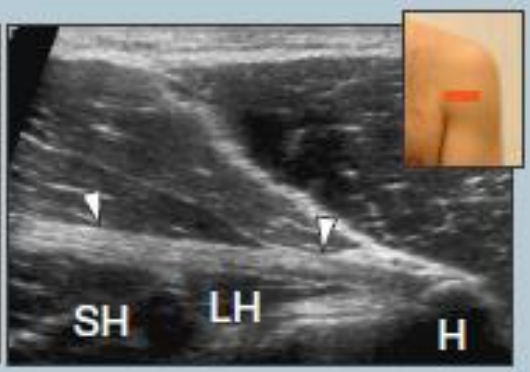
2 long head of the biceps tendon

Place the arm in slight internal rotation (directed towards the contralateral knee) with the elbow flexed 90°, palm up. Start by finding the long biceps tendon in between the greater and lesser tuberosities – Use short and long (more limited utility) axis planes to examine the biceps.





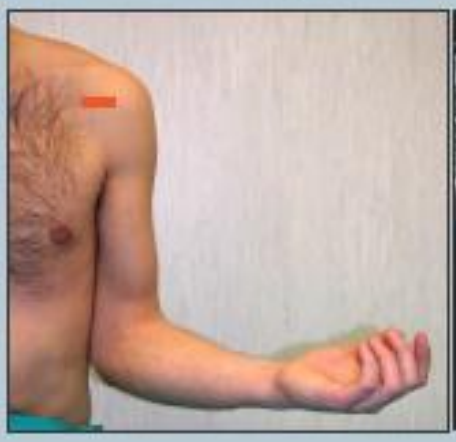
Shift the probe up to examine the biceps in its intraarticular course and down to reach the myotendinous junction (level of the pectoralis major tendon).

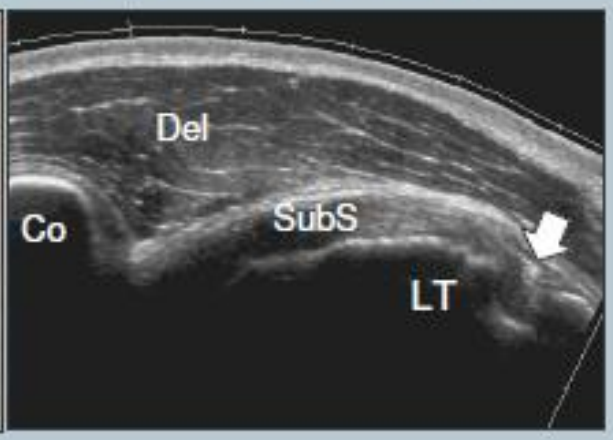


Legend: SubS, subscapularis tendon; SupraS, supraspinatus tendon; Arrow, long head of the biceps tendon; LT, lesser tuberosity; GT, greater tuberosity; SH, short head of the biceps; LH, long head of the biceps; H, humoral shaft; Arrowheads, pectoralis major tendon

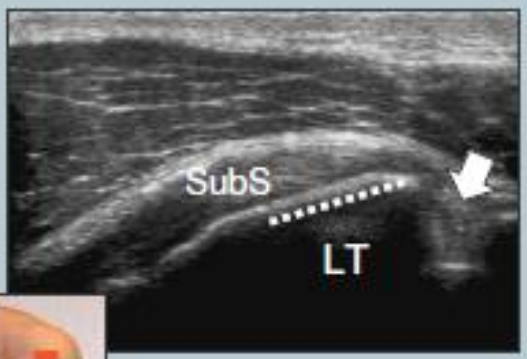
3 subscapularis tendon

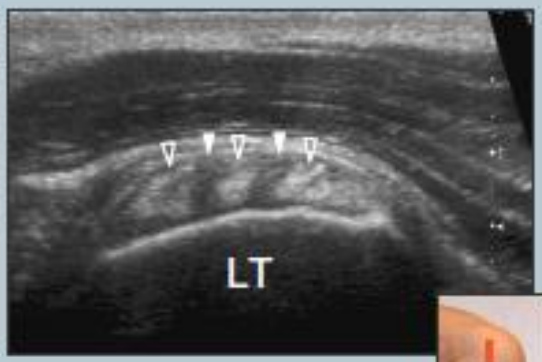
Rotate the arm externally fixing the elbow on the iliac crest to show the subscapularis tendon and its insertion on the lesser tuberosity (slight supination of the hand may be helpful to neutralize the tendency to lift and abduct the elbow from the lateral chest wall).

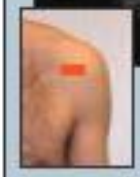


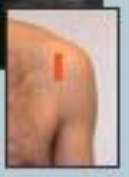


This tendon should be evaluated along its long- (transverse planes) and short- (sagittal planes) axis during passive external and internal rotation with hanging arm. Sweep the transducer up and down over the subscapularis until its full width is demonstrated.





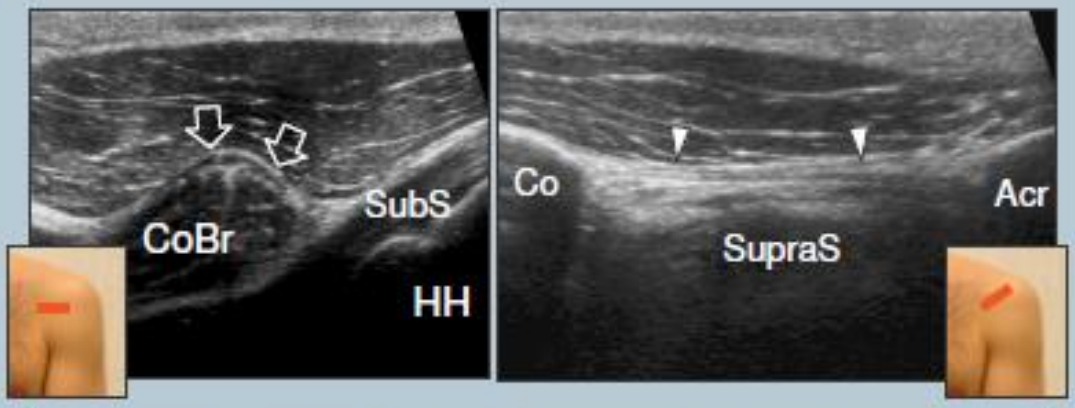




Legend: Arrow, long head of the biceps tendon; dashed line, insertion of the subscapularis tendon; Co, coracoid; Del, deltoid muscle; LT, lesser tuberosity; SubS, subscapularis tendon; void arrowheads, tendon fascicles of the subscapularis; white arrowheads, muscle tissue interposed between tendon fascicles

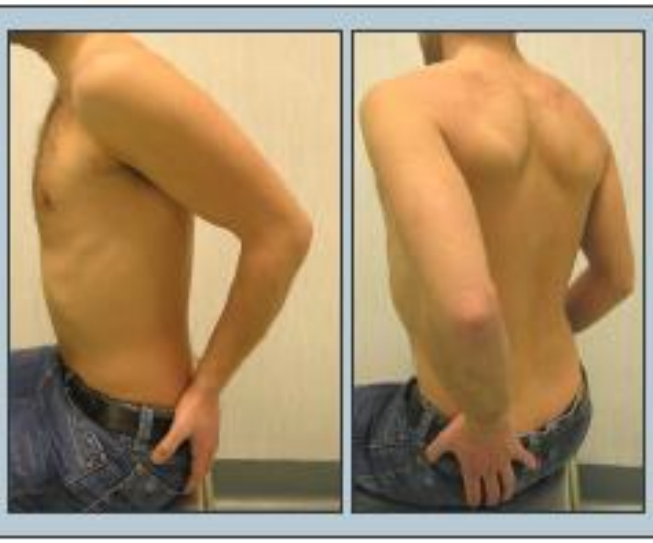
4 anteromedial structures and coracoacromial ligament

Moving the probe toward medial on transverse planes, look at the coracoid process, the coracoacromial ligament (the medial margin of the transducer is on the coracoid process and the lateral edge is shifted upward and laterally towards the acromion), the conjoined tendon and the anterior aspect of the subacromial subdeltoid bursa. Then, check the subscapularis recess and the subcoracoid bursa for effusion. External and internal rotation may also used to demonstrate anteromedial impingement (distance between coracoid process and lesser tuberosity measured in internal rotation).



Legend: Acr, acromion; Arrows, short head of the biceps; Arrowheads, coracoacromial ligament; Co, coracoid; CoBr, coraco-brachialis; HH, humeral head; SubS, subscapularis tendon; SupraS, supraspinatus

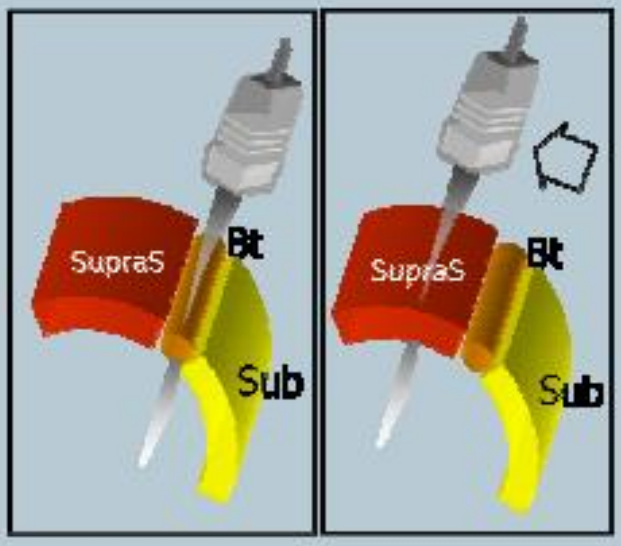
5 supraspinatus tendon: positioning (1)

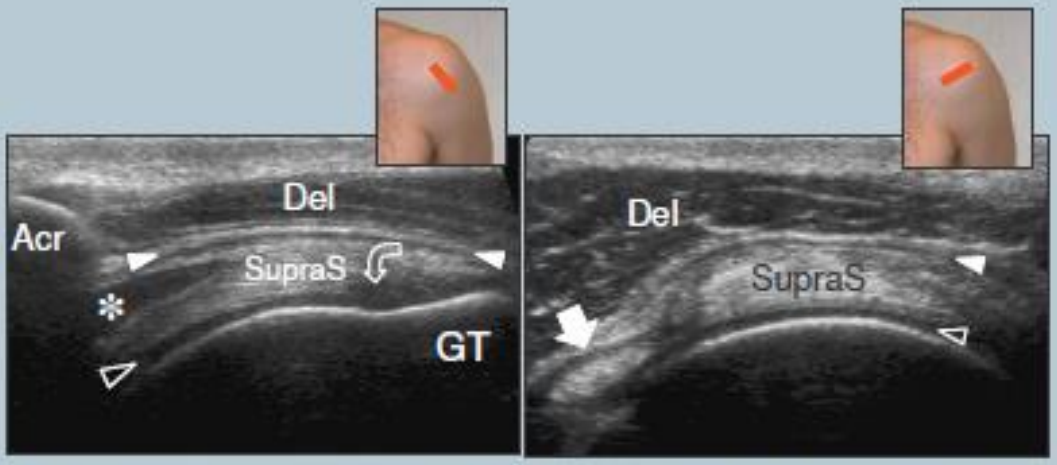


Place the patient's arm posteriorly, placing the palmar side of the hand on the superior aspect of the iliac wing with the elbow flexed and directed posteriorly. The supraspinatus tendon should be evaluated along its long- and short-axis.

5 supraspinatus tendon

Refer to the intraarticular portion of the biceps as a landmark to obtain proper transducer orientation for imaging the supraspinatus. In fact, these tendons run parallel one to the other and the intraarticular portion of the biceps is easy to be recognized due to a more clearly defined fibrillar pattern. One should rotate the transducer until the biceps is depicted as more elongated as possible in the US image. Then, the probe is shifted upward and posteriorly over the supraspinatus without changing its orientation. The resulting image is in axis with the supraspinatus. Between the supraspinatus and the deltoid, the normal subacromial-subdeltoid bursa appears as a thin hypoechoic band.



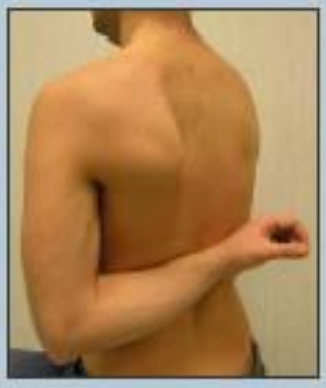


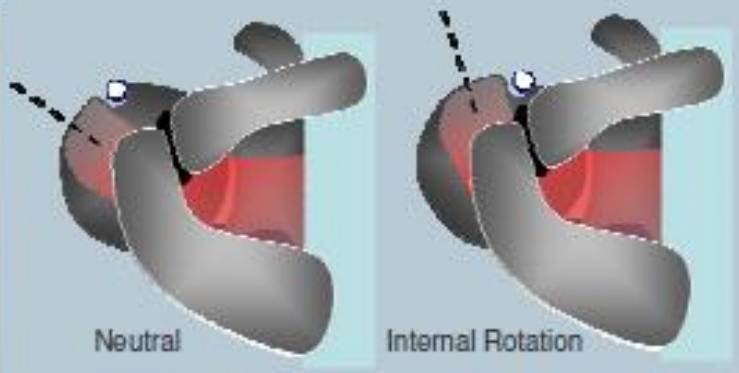
Tilt the transducer gently in the area overlying the tendon insertion to avoid anisotropy. Remember to scan the lateral pouch of the subacromial subdeltoid bursa along the lateral edge of the greater tuberosity. When looking for the supraspinatus on short-axis, the normal cuff must have almost the same thickness from the biceps tendon landmark until 2cm backwards: from this point backwards the tendon seen is the infraspinatus.

Legend: Acr, acromion; asterisk, myotendinous junction; Del, deltoid muscle; GT, greater tuberosity; void arrow-head, articular cartilage; curved arrow, hypoechoic artifact related to anisotropy; straight arrow, long head of the biceps tendon; SupraS, supraspinatus tendon; white arrowhead, subacromial subdeltoid bursa

6 supraspinatus tendon: positioning (2)

Place the dorsum of the hand over the opposite back pocket (forced internal rotation, stress manoeuvre). There should not be any space gap between the elbow and the lateral chest wall. Using this position, the supraspinatus becomes a more anterior structure and the transducer should be oriented almost vertically to be in axis with it. Consider that the tendon fibers are more stretched than in the position described at point -5. This may be possible cause of overestimation of tear size. Due to an excessive internal rotation, the long head of the biceps tendon may be difficult to be visualized in this position.





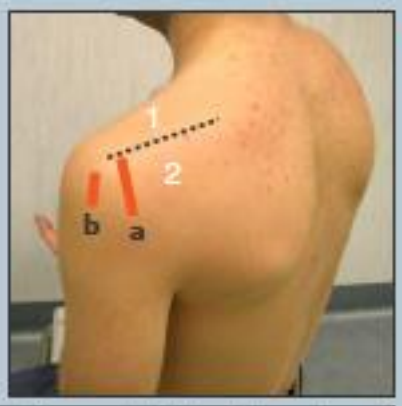
7 subacromial impingement test

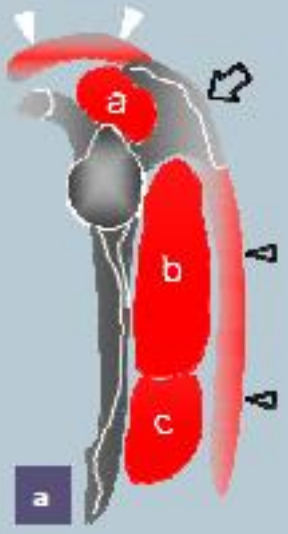
Dynamic assessment of subacromial (antero-superior) impingement can be attempted by placing the probe in the coronal plane with its medial margin at the lateral margin of the acromion. The patient abducts his arm while in internal rotation. With this manoeuvre, the supraspinatus and the bursa can be seen passing deep to the coracoacromial arch.

Legend: Acr, acromion; GT, greater tuberosity; arrows, supraspinatus tendon

8 infraspinatus and teres minor tendons

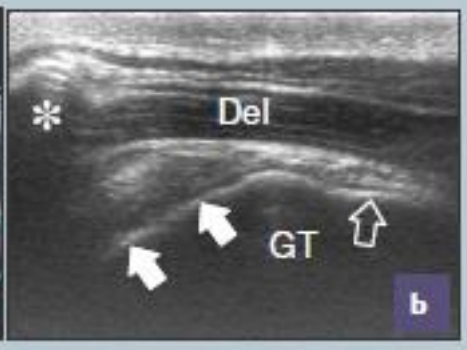
Place the transducer over the posterior aspect of the glenohumeral joint with the arm in the same position described at point-2 (or with the hand on the opposite shoulder) and increase the depth to include the structures of the posterior fossa within the field-of-view of the US image. Use the spine of the scapula as the landmark to distinguish the supraspinous fossa (transducer shifted-up) from the infraspinous fossa (transducer shifted-down) on sagittal planes.





Legend: a, supraspinatus; arrow, spine of the scapula; b, infraspinatus; c, teres minor; dashed line, spine of the scapula; 1, supraspinous fossa; 2, infraspinous fossa; void arrowheads, deltoid; white arrowheads, trapezius

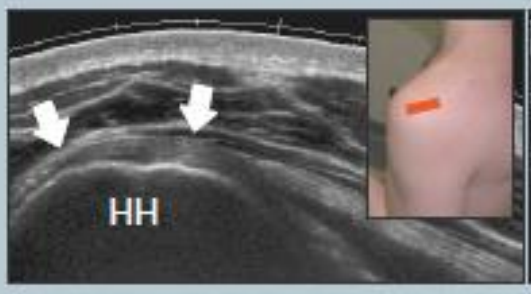
Look at the infraspinatus and teres minor muscles as individual structures filling the infraspinous fossa deep to the deltoid. After scanning these muscles, sweep the transducer toward the greater tuberosity on sagittal planes. The two tendons can be appreciated as individual structures arising from the respective muscles.

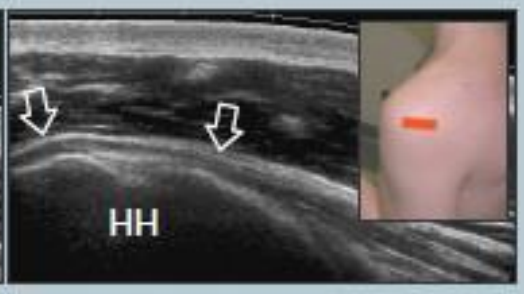


Legend: asterisk, spine of the scapula; dashed line, spine of the scapula; Del, deltoid muscle; GT, greater tuberosity; InfraS, infraspinatus muscle; Tm, teres minor muscle; void arrow, teres minor tendon; white arrows, infraspinatus tendon

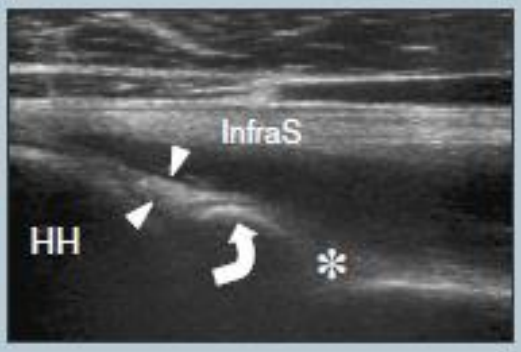
9 posterior structures and posterior glenohumeral joint recess

Examine these tendons separately on their long-axis (transverse planes) during external and internal rotation of the arm (same position as in point-2) by placing the probe over the posterior aspect of the glenohumeral joint.





Look at the posterior labrum-capsular complex and check the posterior recess of the joint for effusion during scanning. In thin subjects the posterior labrum can be clearly seen. Move the transducer medial to the labrum on transverse plane to visualize the spinoglenoid notch. It is often necessary to increase the depth of the field-of-view not to miss this area. A paralabral cyst originating in this area should be sought.

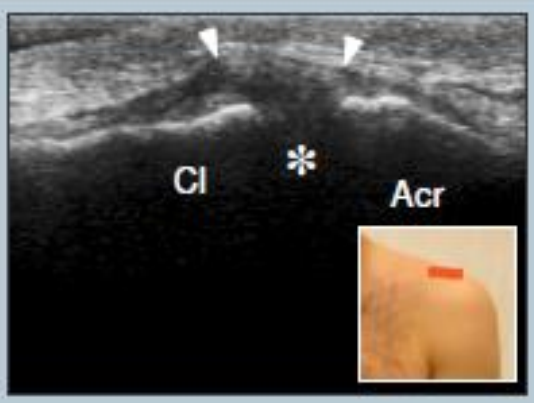


Legend: asterisk, spinoglenoid notch; curved arrow, bony glenoid; HH, humeral head; InfraS, infraspinatus; void arrows, teres minor tendon; white arrows, infraspinatus tendon; white arrowheads, posterior labrum

10 acromioclavicular joint

Place the transducer in the coronal plane over the shoulder to examine the acromioclavicular joint. Sweep the transducer anteriorly and posteriorly over this joint to assess the presence of an os acromiale. Shifting the probe posterior to the acromioclavicular joint, it is possible to assess the status of the supraspinatus muscle.

Legend: Acr, acromion; arrowheads, superior acromioclavicular ligament; asterisk, acromioclavicular joint space; Cl, clavicle



11.12 Ultrasound differential diagnoses and record scan findings

Taken from the publication:

A training, assessment and feedback package for the trainee shoulder sonographer

Michael J Smith[1], Alison Rogers[2], Nazar Amso[2], Julia Kennedy[1], Alison Hall[3] and Peter Mullaney[4]

[1]School of Healthcare Sciences, Cardiff University, Cardiff, UK
[2]Institute for Translation, Innovation, Methodology and Engagement, Cardiff University School of Medicine, Cardiff, UK
[3]Primary Care Sciences, Keele University, Staffordshire, UK
[4]Radiology Department, University Hospital of Wales, Cardiff, UK
Corresponding author: Michael J Smith. ORCiD number (orcid.org/0000-0002-4199-3315), Email: smithmj2@cardiff.ac.uk

Abstract

Diagnostic ultrasound of the shoulder is recognised as being one of the most technically challenging aspects of musculoskeletal ultrasound to master. It has a steep learning curve and makes gaining competency a time-intensive training process for both the trainee and their trainer. This article describes a training, assessment and feedback package developed within the framework of a Consortium for the Accreditation of Sonographic Education approved post-graduate ultrasound course. The package comprises: (i) a shoulder diagnostic ultrasound scan protocol with definition of findings, differential diagnosis and pro forma for recording scan findings, (ii) an assessment form for performance of shoulder diagnostic ultrasound scans with assessment criteria and (iii) a combined performance assessment and scan findings form, for each tissue being imaged. The package has been developed using medical education principles and provides a mechanism for trainees to follow an internationally recognised protocol. Supplementary information includes the differential diagnostic process used by an expert practitioner, which can otherwise be difficult to elicit. The package supports the trainee with recording their findings quickly and consistently and helps the trainee and trainer to explicitly recognise the challenges of scanning different patients or pathologies. It provides a mechanism for trainers to quantify and trainees to evidence their emerging competency. The package detailed in this article is therefore proposed for use in shoulder ultrasound training and its principles could be adapted for other musculoskeletal regions or other ultrasound disciplines.

Keywords: Ultrasound, sonography, training, assessment, shoulder, musculoskeletal

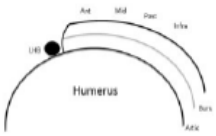
Ultrasound 2015; **23**: 29–41. DOI: 10.1177/1742271X14566067

11.12.1 Section 1b – for differential diagnosis

Section 1b: Definition of scan findings and differential diagnosis (i.e. how to interpret the images + helpful tips)

Step *(ESSR)*	Image/tissue	Definition of findings/differential diagnosis (+ *helpful tips*)	Area of controversy or advanced/ confirmatory technique
1*(2)*	Transverse view of biceps tendon	• Tendon present and normal – Visible as hyper-echoic structure; characterised by anisotropy. • Tendon torn (partial) – Hyper-echoic structure located but abrupt change in cross-sectional shape or size; location of change reported in relation to bicipital groove. • Tendon torn (complete) – Hyper-echoic structure cannot be located (*requires confirmation in two planes*). • Tendon displaced – Present, but not in inter-tubercular groove; dislocates.	• Tendinopathic change – Thickening or thinning of tendon; hypo-echoic, irregular signal/fibre appearance. • Tenosynovitis –thickening of synovial sheath and/or increased fluid.
2*(2)*	Longitudinal view of biceps tendon	• As per step 1.	• As per step 1.
3	Dynamic transverse assessment of biceps tendon	• Tendon stable in inter-tubercular groove. • Tendon displaced/subluxes, during GHJ internal/external rotation.	
4*(3)*	Longitudinal view of Subscapularis	• Tendon present and normal – Visible by characteristic shape and attachment to lesser tuberosity. • Tendon torn (partial) – Hypo-echoic region which does not extend through full tendon thickness (confirm in two planes), +/– fluid or tissue in-fill. • Tendon torn (full) – Hypo-echoic region which does extend through full tendon thickness (confirm in two planes), +/– fluid or tissue in-fill. • Tendon torn (complete) – Hypo-echoic region which extends through full tendon thickness and width +/– retraction (confirm in two planes), +/– fluid or tissue in-fill.	• Tendinopathic change – Thickening or thinning of tendon; hypo-echoic, irregular signal/fibre appearance; calcific deposits.
5*(3)*	Transverse view of Subscapularis	• Tendon present and normal – Visible by characteristic shape and fibrillar arrangement (*note fibrillar arrangement is a normal finding and should not be confused with tendinopathic change or tear*). • Tendon torn (partial, full or complete) – As per step 4 (above); confirm tear in two planes.	• Tendinopathic change – Thickening or thinning of tendon; hypo-echoic, irregular signal/fibre appearance; calcific deposits.
6*(10)*	Acromioclavicular joint	• Joint margins smooth and normal. • Osteophytic – Irregular joint margins with bony outgrowths; graded according to size and irregularity of outgrowths.	• Does the patient report reproduction of their symptoms upon scanning or pressure from probe?

Continued

Step *(ESSR)*	Image/tissue	Definition of findings/differential diagnosis (+ *helpful tips*)	Area of controversy or advanced/ confirmatory technique
		• Synovitis/Inflammation – Capsular hypertrophy; localised vascular signal present (*independent of patient breathing/ talking*).	
7*(5)*	Transverse view of Supraspinatus	• Tendon present and normal – Visible by characteristic appearance:  • Tendon torn (partial) – Hypo-echoic region which does not extend through full tendon thickness (confirm in two planes), +/− fluid or tissue in-fill; identify bursal/intra-substance/articular side. • Tendon torn (full) – Hypo-echoic region which does extend through full tendon thickness (confirm in two planes), +/− fluid or tissue in-fill; identify bursal/intra-substance/articular side. • Tendon torn (complete) – Hypo-echoic region which extends through full tendon thickness and width +/− retraction (confirm in two planes), +/− fluid or tissue in-fill.	• Tendinopathic change – Thickening or thinning of tendon; hypo-echoic, irregular signal/fibre appearance; calcific deposits. • Bursa – Fluid present (hypo-echoic)/ thickening; debris present (mixed echogenicity).
8*(5)*	Longitudinal view of Supraspinatus	• Tendon present and normal – Visible by characteristic ''birds beak'' shape and attachment to greater tuberosity. • Tendon torn (partial/full thickness/complete) – As per step 7; confirm in two planes.	• Tendinopathic change – thickening or thinning of tendon; hypo-echoic, irregular signal/fibre appearance; calcific deposits. • Bursa – Fluid present (hypo-echoic)/ thickening; debris present (mixed echogenicity).
11*(7)*	Subacromial impingement test + visualisation of tissue in subacromial region		• Is there smooth passage of subacromial tissues under the acromion? • Does the patient report reproduction of their symptoms when moving? • From anterior to posterior subacromial region, is acromio-humeral distance preserved/tendon tissue present throughout the observable region?
9*(8)*	Transverse view of Infraspinatus (and Teres Minor)	• Muscle(s) through to tendon(s) present and normal (*visible by characteristic shape*). • Tendon torn (partial) – Hypo-echoic region which does not extend through full tendon thickness (confirm in two planes), +/− fluid or tissue in-fill. • Tendon torn (full) – Hypo-echoic region which does extend through full tendon thickness (confirm in two planes), +/− fluid or tissue in-fill. • Tendon torn (complete) – Hypo-echoic region which extends through full tendon thickness and width +/− retraction (confirm in two planes), +/− fluid or tissue in-fill.	• Estimate if muscle thickness is within normal limits; loss of marbled muscle appearance for fatty infiltration. • Tendinopathic change – Thickening or thinning of tendon; hypo-echoic, irregular signal/fibre appearance; calcific deposits.
10 *(9)*	Longitudinal view of Infraspinatus (and Teres Minor) Posterior glenohumeral joint recess	• Tendon(s) present and normal – Visible by characteristic shape and attachment to greater tuberosity; identify posterior glenoid. • Tear (partial/full/complete)? • Joint effusion or labral cyst?	• Tendinopathic change – Thickening or thinning of tendon; hypo-echoic, irregular signal/fibre appearance; calcific deposits.

Note: *EESR* indicates the corresponding step in the ESSR guidelines. Where a structure (or ''change'' in a structure) has been identified as a pathological finding and/or potential cause of pain, then comparison with the structure of interest on the contralateral side should be performed. It is acknowledged that asymmetry can be a ''normal'' finding; nonetheless, comparison with the contralateral side can provide useful confirmatory information.

11.12.2 Section 1c – for scan findings recording

Section 1c: Shoulder diagnostic ultrasound scan – Scan findings pro forma

Patient ID / code: **Date and time:**

Side: *L* *R* **Name of trainee / trainer:**

Biceps Tendon (Steps 1,2,3):

Normal *Tear* Partial / Complete *Displaced / dislocating* *Tendinopathic change* *Tenosynovitis*

Details:

Subscapularis (Steps 4,5):

Normal *Tear: Partial* *Full* *Complete* *Tendinopathic change / calcific deposit*

Details:

ACJ (Step 6):

Normal *Osteophytic mild / mod / severe* *Synovitis / Inflammation*

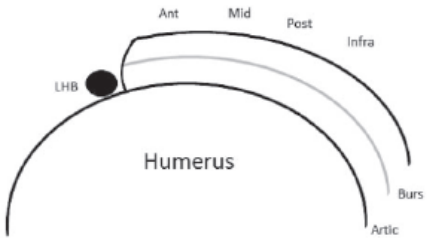
Details:

Supraspinatus (Steps 7,8 & 11):

Normal *Tear (X) Partial / Full / Complete* *Tendinopathic change / calcific deposit (O)*

Bursa Fluid / thickening / debris

Details:
(incl smooth passage of tissues on dynamic testing and whether acromio-humeral distance is preserved / tendon tissue is present)



Infraspinatus (and Teres Minor) and posterior glenohumeral joint recess (Steps 9,10):

Normal *Tear: Partial* *Full* *Complete* *Joint effusion Labral cyst* *Tendinopathic change / calcific deposit*

Details:

11.13 Demographics and clinical data of ultrasound inter-rater reliability study subjects

Table 11-11: Demographics and clinical data of ultrasound inter-rater reliability study subjects

Pt code	Gender 1=M, 2=F	Age (years)	Sx side 1=L, 2=R	Referral source (coding below)	Onset (coding below)	Chronicity (months)	Prev episodes (coding below)
F4589	2	73	2	1	1	5	0
F4793	2	25	2	1	2	8	0
F1953	2	82	2	1	2	36	1
M3535	1	25	2	1	1	6	0
F3577	2	65	2	1	3	8	0
F3317	2	63	2	4	2	1	0
M6133	1	51	1	1	2	12	0
F7185	2	47	1	1	2	9	0
F2815	2	56	1	1	2	9	0
F5545	2	35	2	1	3	60	0
M0833	1	63	2	1	2	216	1
F1194	2	28	2	1	4	22	0
F4814	2	45	1	1	2	24	0
M5213	1	52	1	1	1	4	0
M3023	1	69	2	1	2	12	0
F7914	2	57	1	1	2	9	0
F8687	2	76	1	2	5	36	0
F1706	2	59	2	1	5	18	0
M9997	1	42	1	2	n/r	30	1
M7087	1	57	2	1	2	36	0
F7537	2	47	2	1	3	3	0
M2353	1	54	2	1	2	8	0
M3933	1	57	2	1	2	24	0
F0842	2	57	2	1	3	12	0
F5937	2	44	2	1	2	96	0
M8858	1	61	2	1	2	3	0
F0962	2	82	2	1	2	6	0
M4929	1	77	2	1	2	7	0
F3183	2	76	1	1	2	24	0
M0454	1	40	1	1	1	8	0
M5901	1	45	2	1	2	15	0
M0902	1	50	1	1	2	24	0
F8015	2	52	2	n/r	2	18	0
M5381	1	21	1	4	1	2	0
F9032	2	58	2	1	2	24	0

Key: M = male, F = female, L = left, R = right, Referral source: 1 = GP; 2 = Rheumatology; 3 = Orthopaedics; 4 = Trauma/ fracture clinic, Onset: 1=Trauma; 2=Insidious; 3=combination of 1 & 2; 4=overuse; 5 = Rheumatoid Arthritis, Previous episodes: 0 = none; 1 = multiple, n/r = not recorded

11.14 Structural pathology findings from ultrasound inter-rater reliability study

Table 11-12: Structural pathology findings from ultrasound inter-rater reliability study

Rater	MS	PM	MS	PM	MS	PM	MS	PM	MS	PM	MS	PM
Pt code	Cuff path 1=Y, 2=N	Cuff path 1=Y, 2=N	Tendino pathic 1=Y, 2=N	Tendino pathic 1=Y, 2=N	Calcific deposit 1=Y, 2=N	Calcific deposit 1=Y, 2=N	Bursitis 1=Y, 2=N	Bursitis 1=Y, 2=N	PTT 1=Y, 2=N	PTT 1=Y, 2=N	FTT 1=Y, 2=N	FTT 1=Y, 2=N
F4589	1	1	1	1	2	2	2	2	2	2	1	2
F4793	2	2	2	2	2	2	1	1	2	2	2	2
F1953	1	2	1	2	2	2	2	2	2	2	2	2
M3535	2	2	2	2	2	2	1	2	2	2	2	2
F3577	1	2	1	2	2	2	1	1	2	2	2	2
F3317	1	2	1	2	2	2	2	2	2	2	2	2
M6133	2	2	2	2	2	2	2	2	2	2	2	2
F7185	2	1	2	1	2	2	1	2	2	2	2	2
F2815	2	2	2	2	2	2	2	2	2	2	2	2
F5545	2	2	2	2	2	2	1	1	2	2	2	2
M0833	1	1	1	1	2	2	1	1	2	2	1	1
F1194	1	1	1	1	2	2	1	n/r	2	2	2	2
F4814	2	1	2	1	2	2	2	n/r	2	2	2	2
M5213	1	1	2	1	2	2	2	2	2	2	1	1
M3023	1	2	1	2	2	2	1	1	2	2	2	2
F7914	1	1	1	1	2	2	1	n/r	2	2	2	2
F8687	1	1	1	1	1	2	2	n/r	2	1	2	2
F1706	1	1	1	1	2	2	2	n/r	2	2	2	2
M9997	2	1	2	1	2	2	2	n/r	2	2	2	2
M7087	1	2	1	2	2	2	2	2	2	2	2	2
F7537	2	1	2	1	2	2	1	n/r	2	2	2	2
M2353	1	1	1	1	1	2	2	n/r	2	2	2	2
M3933	1	1	1	2	2	2	2	n/r	2	2	1	1

F0842	1	1	2	1	1	1	n/r	n/r	2	2	2	2
F5937	1	2	1	2	1	2	1	n/r	2	2	2	2
M8858	1	1	1	1	n/r	1	1	2	2	2	2	2
F0962	1	1	1	1	2	2	1	1	2	1	2	2
M4929	1	1	2	1	2	2	1	n/r	1	1	2	2
F3183	1	1	1	1	2	2	n/r	n/r	2	2	2	2
M0454	n/r	1	n/r	1	n/r	2	2	2	n/r	2	n/r	2
M5901	2	2	2	2	2	2	1	n/r	2	2	2	2
M0902	1	1	1	1	2	2	1	n/r	2	2	2	2
F8015	1	1	1	1	2	2	1	2	2	2	2	2
M5381	2	2	2	2	2	2	1	n/r	2	2	2	2
F9032	1	1	1	1	1	2	1	2	2	2	2	2

Key: Y = yes, N = no, Cuff path = Rotator cuff pathology, PTT = Partial Thickness Tear; FTT = Full Thickness Tear, n/r = not recorded
Note: Pathology findings are collapsed for all rotator cuff tendons

11.15 Preliminary version of proforma for collecting patient treatment information

Please place patient sticky here

1. Date of **first** patient appointment (i.e. assessment):

2. Name and grade of clinician:

Treatment delivered		
"Modality"	**Delivered during appointment? If so, the approximate number of minutes you spent with the patient doing this**	**Given as a home exercise programme?**
Advice (including explanation about pathology, etc)		
Manual therapy (Glenohumeral mobilisation; spinal mobilisation)		
Exercise therapy: Glenohumeral stability – including rotator cuff strengthening		
Exercise therapy: Scapulothoracic stability		
Exercise therapy: Glenohumeral mobility / range of movement		
Taping: what technique?		
Electrotherapy: what modality?		
Stretching: Anterior shoulder / pec major / pec minor		
Stretching: Posterior shoulder / sleeper stretch		
Other:		

Any other comments?

11.16 Final version of proforma for collecting patient treatment information

Patient treatment information

Dear Colleague,

Your patient is participating in the Shoulder Prognosis Study that xxx and I are leading here in the Department.

As part of the study, please use this proforma to record the content of **each** treatment session (including home exercises) with your patient. This information will only be used in secondary analysis and will be anonymised.

Treatment is entirely clinician directed – i.e. you treat the patient as many times and using whatever modalities you feel is appropriate. From a clinician perspective the study is NOT about making a judgement on whether you use the "right" or "wrong" modalities or whether many or few of your patients improve. The study is only concerned with whether specific characteristics can predict which of the patients we see can be helped with Physiotherapy.

Thank you very much for your support.

Best wishes
Mike

Email: Smithmj2@cf.ac.uk Telephone: (029) 2068 7927

Date of **first** patient appointment (i.e. assessment)		*Please place patient sticky here*
Name of clinician		
Clinician D&T code (and therefore grade)		

Clinic-based treatment – please **circle** all that apply

Education and advice	Explanation about pathology	Avoiding certain movements / activities	Postural / ergonomic / functional advice	Other…..
Manual therapy	Glenohumeral / ACJ mobilisation / manipulation	Spinal mobilisation / manipulation	Soft tissue massage / release, e.g. trigger point, hold-relax	
Exercise therapy	Shoulder **ROM** e.g. pendulum, active-assisted	Shoulder **flexibility** e.g. anterior or posterior stretches	Shoulder **strengthening** incl rotator cuff	**Scapular stabiliser** control and/or strength	**Core** stability exercises
Taping	Inhibitory / facilitatory e.g. upper / lower traps	Postural e.g. scapular or thoracic	Relocation / support e.g. “Adidas” technique	
Electrotherapy	Ultrasound	TENS	Interferential	Short wave diathermy	
Other interventions	Heat	Cold	Acupuncture	Hydrotherapy	Injection

Additional details (if required) regarding treatment:

Home exercise programme – please **circle** all that apply

Exercise therapy	Shoulder **ROM** e.g. pendulum, active-assisted	Shoulder **flexibility** e.g. anterior or posterior stretches	Shoulder **strengthening** incl rotator cuff	**Scapular stabiliser** control and/or strength	**Core** stability exercises

Any other comments?

12 Appendix V: Methods chapter

Prognostic cohort study subjects:
- Study invitation pack
- Order of testing

C1. Patient reported measures: Pain
- Visual Analogue Scale

C2. Patient reported measures: 4DSQ
- 4DSQ: Patient to complete
- 4DSQ: Scoring and interpretation

C3. Patient reported measures: Function / disability
- SPADI

Outcome
- OSS: Questionnaire
- OSS: Scoring

Treatment
- Patient data collection form

D3. Clinical measures: Scapular movement and control
- SDT assessment criteria

E1. Structural pathology via imaging
- Specific criteria relating to the sonographic appearance of cuff tendinopathy and bursitis.
- Philips CX-50

12.1 Study invitation pack

This comprised a covering letter for each site (UHW or WHI), a patient self-screening form and patient information sheets 1 and 2

12.1.1 Covering letter to patient for UHW site

GIG CYMRU NHS WALES
Bwrdd Iechyd Prifysgol Caerdydd a'r Fro
Cardiff and Vale University Health Board

Ysbyty Athrofaol Cymru

University Hospital of Wales

Heath Park, Cardiff, CF14 4XW

Parc Y Mynydd Bychan, Caerdydd, CF14 4XW

Eich cyf/Your ref

Ein cyf/Our ref

Welsh Health Telephone Network 1872

Dear ,

We are contacting you as you have been referred for Physiotherapy at the University Hospital Wales in the Heath.

The Physiotherapy Department is collaborating with Cardiff University to look at how information from shoulder pain patients can be used to predict which patients will respond to Physiotherapy. We would therefore like to invite you to participate in the study.

The data collection will involve performing an ultrasound scan of your shoulder, measuring your shoulder strength and videoing your shoulder movements. The data collection will occur immediately prior to your first Physiotherapy appointment **so there will be no additional travelling or attendance and will last no longer than 45 minutes**. Your Physiotherapy assessment and treatment will occur as normal.

We would be grateful if you would read the enclosed information sheet, please. The one entitled "Are your shoulder symptoms suitable for our study?" will help you to identify if you are eligible to be included in the study.

If you are interested in participating in the study then – once you have received your letter from the Physiotherapy Department asking you to call to make an appointment – please telephone the Physiotherapy Department on (029) 2074 5884, tell the receptionist that you are in the "Mike Smith Shoulder Study" and make your first appointment for either a Monday or a Tuesday.

If you have any questions, queries or concerns about the study then please contact the researcher (Mike Smith) via email (smithmj2@cf.ac.uk) or phone / text (07531 711 508). One of the research team will contact you prior to your first appointment to explain the study and answer any questions that you may have.

Thank you for considering participating in this study.

Yours sincerely,

Adrian Broad
Strategic Lead Outpatient Physiotherapy

12.1.2 Covering letter to patient for WHI site

GIG CYMRU NHS WALES
Bwrdd Iechyd Prifysgol Caerdydd a'r Fro
Cardiff and Vale University Health Board

Ysbyty Athrofaol Cymru

University Hospital of Wales

Heath Park, Cardiff, CF14 4XW

Parc Y Mynydd Bychan, Caerdydd, CF14 4XW

Eich cyf/Your ref

Ein cyf/Our ref

Welsh Health Telephone Network 1872

Dear ,

We are contacting you as you have been referred for Physiotherapy at Whitchurch hospital.

The Physiotherapy Department is collaborating with Cardiff University to look at how information from shoulder pain patients can be used to predict which patients will respond to Physiotherapy. We would therefore like to invite you to participate in the study.

The data collection will involve performing an ultrasound scan of your shoulder, measuring your shoulder strength and videoing your shoulder movements. The data collection will occur immediately prior to your first Physiotherapy appointment **so there will be no additional travelling or attendance and will last no longer than 45 minutes**. Your Physiotherapy assessment and treatment will occur as normal.

We would be grateful if you would read the enclosed information sheet, please. The one entitled "Are your shoulder symptoms suitable for our study?" will help you to identify if you are eligible to be included in the study.

If you are interested in participating in the study then – once you have received your letter from the Physiotherapy Department asking you to call to make an appointment – please telephone the Physiotherapy Department on (029) 2033 6345, tell the receptionist that you are in the "Mike Smith Shoulder Study" and make your first appointment for a Thursday.

If you have any questions, queries or concerns about the study then please contact the researcher (Mike Smith) via email (smithmj2@cf.ac.uk) or phone / text (07531 711 508). One of the research team will contact you prior to your first appointment to explain the study and answer any questions that you may have.

Thank you for considering participating in this study.

Yours sincerely,

Adrian Broad
Strategic Lead Outpatient Physiotherapy

12.1.3 Patient self-screening form

Are your shoulder symptoms suitable for our study?

Your Name	**Today's Date**	
Thank you for helping us with our shoulder study. Please complete the questions below to see whether you have the type of shoulder problem that we are investigating in our study. **The questions are about the shoulder problem for which you have been referred to physiotherapy. The questions are not about your other shoulder.** Please tick one box to indicate your answer for each question.		
1. Have you been referred to physiotherapy for a problem with your shoulder?	Yes ☐	No ☐
2. Is your shoulder pain worse when you move into a particular position or do specific tasks?	Yes ☐	No ☐

I answered "Yes" to questions 1 and 2. What do I do now?
Please continue and answer questions 3 to 8 on the other side of this sheet of paper
I answered "No" to question 1 or 2. What do I do now?
You do not have the type of shoulder problem that we are investigating in our study. Please do not answer any further questions. Please attend your first physiotherapy appointment as usual. Your treatment will be unaffected. Thank you for your time and interest in the study.

3. Have you had an operation on your shoulder in the last five years?	Yes ☐	No ☐
4. Have you broken or fractured one of the bones in your shoulder in the last five years?	Yes ☐	No ☐
5. Has your shoulder "popped out of joint" or dislocated in the last five years?	Yes ☐	No ☐
6. Is stiffness in your shoulder much more of a problem for you than *pain* in your shoulder?	Yes ☐	No ☐
7. Did your shoulder pain come on suddenly following an accident that happened in the last 4 weeks?	Yes ☐	No ☐
8. Do your shoulder symptoms get worse when you move *your neck* rather than your shoulder in a particular position?	Yes ☐	No ☐

I answered "No" to questions 3 to 8. What do I do now?

You have the type of shoulder symptoms we are investigating in our study.

Please open the envelope and read the information sheets. Please wait until close to the date of your first appointment before filling in the questionnaire booklet.

Please bring the completed questionnaire booklet to your first appointment

I answered "Yes" to one or more questions from 3 to 8. What do I do now?

You do not have the type of shoulder problem that we are investigating in our study. Please do not answer any further questions. Please attend your first physiotherapy appointment as usual. Your treatment will be unaffected. Thank you for your time and interest in the study.

12.1.4 Patient information sheet 1

PATIENT INFORMATION SHEET

Assessment of joint function in patients with joint problems using three dimensional motion analysis techniques

Version 7 21/02/2013

Part one

You are being invited to take part in a research study with Cardiff University's Arthritis Research UK Biomechanics and Bioengineering Centre. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. One of our team will go through the information sheet with you. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to participate. Part 1 tells you about the purpose of this study and what will happen to you if you take part. Part 2 gives you more detailed information about the conduct of the study.

What is the purpose of this trial?
The aim of the trial is to investigate the function of joints for people with joint problems and people with healthy joints. The data can be used to develop new treatments, improve the design of joint replacements, improve rehabilitation and improve the way that motion is analysed clinically.
The study is designed to examine the effects of joint problems and any subsequent operation or other treatment (where appropriate), on the joints ability to perform daily tasks (such as walking, lifting a cup etc).

Do I have to take part?
It is up to you to whether or not to take part. If you do decide to take part you will be given this information sheet to keep and after you have had enough time to read through it, be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time or without giving a reason. A decision not to take part or to withdraw at any time will not affect the standard of care you receive. Should you decide not to take part, you do not have to provide a reason for this decision.

What will happen to me if I take part?
You have been asked to take part in this as you have a problem with your joint that we are interested in looking at with this technique. It will allow us further insight into the nature of joint function and pain that people with your joint problem encounter. You may also been asked to take part so we can examine a non affected joint so we can compare it to the joint problem.

If you wish to take part you will assessed either in the Cardiff University School of Engineering, Human Motion Analysis Laboratory or in the Cardiff University School of Healthcare studies (SOHCS) Research Centre for Clinical Kinaesiology (RCCK) or in the relevant clinical settings. The number of times we would ask you to attend would depend on the joint problem; we will discuss this with you when going through this information sheet.

Data will be kept securely for a minimum of 15 years in accordance with good research practice and data protection regulations imposed by Cardiff University in accordance with the Data Protection Act 1998. All data obtained during the study will remain confidential. Access to data will only be

available to the investigators attached to the Arthritis Research UK Biomechanics and Bioengineering Centre at Cardiff University.

If new information becomes available, we may invite you to take part in a follow-up study in the future, please indicate on the consent sheet if you do not mind us contacting you.

What will I have to do?
Before your first assessment you will be asked to sign a patient consent form which includes the following clause: I understand that I may withdraw from the study at any time without it affecting my ongoing treatment in any way.

At the beginning of your visit, we will explain the study in full and ask for your consent, bearing in mind that you are free to withdraw at any time.

We will ask you to complete questionnaires that will ask you questions about how the problem affects your activities of daily living.

Prior to the start of the assessment, you may be asked to change into appropriate clothing depending on the joint we want to examine (for example shorts for knee, well fitting vest, sports bra, swimming costume, vest or special apron that leave your chest covered and back bare for shoulder and spine, etc). This process will be conducted with the upmost professionalism and a screened off area is provided for changing. During laboratory sessions, access to the laboratory is limited and a sign is placed on the door advising other staff not to enter whilst the trial is in progress.

You will have a number of very light polystyrene or cork round markers attached to the skin and the locations of the markers will be dependent on the joint type under examination.

You will be asked to perform a range of activities of daily living as appropriate (such as walking, standing, climbing stairs, combing hair, taking hand to mouth). You will be free to stop for a break at any time. The position of the markers on the skin will provide a series of recordings by using cameras that record the position of the markers.

You may be asked to perform some of these tasks in a virtual reality environment on a special treadmill. The treadmill is set at floor level and can rotate in an upward/downward direction, move sideways or judder. These movements are controlled and you will be informed if the treadmill will move or rotate. There will be a screen in front of you that can display a variety of images to give the impression of walking on a forest path or kicking a ball, for example. You will be asked to wear a special harness whilst on the treadmill to prevent you from falling in case you stumble.

When appropriate to the joint under study, muscle activity, muscle function and joint strength may also be determined during these sessions. This will involve placement of electromyography (EMG) electrodes onto the surface of the skin to record muscle activity during joint movement. The locations of the electrodes will be dependent on the muscle groups under examination. Particularly hairy skin may sometimes need a small patch shaving for the sensors to attach (approximately 2x2cm). In order to determine muscle function electrical muscle stimulation will be used. This involves placing similar electrodes to the EMG on your skin. During certain movements a small stimulus will be applied via the electrode on your skin, this will make your muscle contract more and change your movement slightly. This may cause a strange sensation but will not cause any pain.

Throughout the sessions your joint movement will be recorded using standard audiovisual equipment. We may ask if we can cover any identifying tattoos or birthmarks with a bandage. Your

face and any identifying tattoos or birthmarks not covered will be digitally masked from these files so that nobody can identify you from the videos. All data files, including audiovisual files will be stored in encrypted folders on Cardiff University password protected computers. Cardiff University and NHS members of staff who are directly involved with the study will have access to the files.

For studies investigating back pain we will ask you to perform a selection of tasks consisting of everyday functional tasks such as bending, stretching, lifting a cup from a table and finding the best position to sit and stand in. Spinal movements and how muscles work when walking may also be assessed whilst you are walking on a treadmill at different speeds and different inclinations.

We will be looking at which targeted exercise treatments using different instructions are the most beneficial for patients with back pain. These will be compared to treatments currently being used such as general advice and general group exercises.

For studies investigating patient with joint osteoarthritis we will determine the best muscle strengthening programmes including how often and how much exercise a patient needs to get an improvement in their joint pain.

For studies investigating wrist osteoarthritis, we will ask you to have a series of measurements and clinical tests performed on both of your wrists, these will include assessing your grip, range of motion and muscle strength.

For studies investigating shoulder pain, we will ask you to perform a series of actions to measure the movement of your shoulder.

Regular rest and toilet breaks will be provided as often as you need them to assure maximum comfort.

Are there any risks in participating in this trial?
The measurements taken during the trial involve the placement of very light polystyrene or cork round markers onto the skin or EMG electrodes in various places of the body depending on what joint we will be examining. The markers/electrodes are placed with sticky tape which may cause some mild discomfort when it is being removed, similar to removing a small sticking plaster.

Are there any benefits in participating in this trial?
We hope to be able to better understand how joint problems affect the motion of the joint. There is no intended clinical benefit to the participant from taking part in the study. The information we get from this study may help us to provide future patients who have joint disease or injury with improved treatment options.

If the information in Part 1 has interested you and you are considering participation, please read the additional information in Part 2 before making a decision.

Part Two

What if new information becomes available?
Sometimes during the course of a research project, new information becomes available about the investigation. If you decide to withdraw, it will not affect your any care in the NHS. If you decide to continue, you will be asked to sign an updated consent form.

What will happen if I do not want to carry on with the study?
If you withdraw from the study, we will erase all identifiable material, but we will need to use the data collected up to your withdrawal.

What if something goes wrong?
If you are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed due to someone's negligence, then you may have grounds for a legal action but you may have to pay for it. Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you have been approached or treated during the course of this study, the normal National Health Service complaints mechanisms should be available to you.

Will my taking part in this study be kept confidential?
All information which is collected about you during the course of the research will be kept strictly confidential. Any information about you which leaves the Cardiff University or the University Hospital of Wales will have your name and address removed so that you cannot be recognised from it. We may share information (including related medical findings such as radiological images) with external collaborators but all this information will contain no identifiable information about you.

Will my GP be informed of my involvement in the study?
With your permission, we will send a letter to your General Practitioner informing him or her of your involvement in the study.

What will happen to the results of the research study?
The measurements taken will provide information about the movement of your joint. The results of the study will be presented at meetings of orthopaedic surgeons, clinical scientists, physiotherapists and engineers, and if accepted, published in medical and engineering journals. If interested, a copy of the published article can be made available to you. You will not be identified in any report/publication.

Who is organising and funding the research?
Research staff at the Arthritis Research UK Biomechanics and Bioengineering Centre at Cardiff University and Consultant Orthopaedic Surgeons at the University Hospital of Wales are carrying out the study. The study is part of the Arthritis Research UK Biomechanics and Bioengineering Centre at Cardiff University; it is not funded by commercial sources and runs alongside research in the Cardiff Gait and Motion Analysis Laboratory at Cardiff University School of Engineering and Research Centre for Clinical Kinaesiology at Cardiff University School of Healthcare Studies.

Who has reviewed the study?
This study has been reviewed by the Research Ethics Committee (REC) for Wales.

What if I wish to lodge a complaint?
If you wish to make a minor complaint regarding the way you were approached or treated during the trial, please contact the Arthritis Research UK Biomechanics and Bioengineering Centre Research

Coordinator at the contact details below or you can contact the Cardiff University Research Governance Team on 029 208 79277.

Contact for further information

Research Coordinator
Arthritis Research UK Biomechanics and Bioengineering Centre
Cardiff School of Biosciences
Cardiff University
Cardiff
CF10 3AX
Tel: 029 2087 5419
Email: Robertshc@cf.ac.uk or Longmanaj@cf.ac.uk

This completes Part 2. Thank you for reading this information sheet.

If you agree to take part in this study then you will be given a copy of the information sheet and a signed consent form to keep.

12.1.5 Patient information sheet 2

See appendix 4 (Pre methods) “Shoulder ultrasound reproducibility study patient information sheet”

12.1.6 Patient consent form 1

PATIENT CONSENT FORM

Assessment of joint function in patients with joint problems using three dimensional motion analysis techniques

Study Number
Patient Identification Number for this trial:

You DO NOT have to sign this document. Please DO NOT sign this document unless you fully understand it. If there is ANYTHING which you do not understand please do not hesitate to ask for a full explanation.

To confirm agreement with each of the statements below, please initial each box and amend as applicable:

1. I confirm that I have read and understand the information sheet dated 21/02/2013 (Version 7) for the above study and have had the opportunity to ask questions. ☐

2. I understand that my participation in the study is voluntary and that I am free to withdraw at any time, without giving any reason, and without my medical care or legal rights being affected. ☐

3. You may/may not contact me in the future to take part in other research projects or surveys. ☐

4. I do/do not agree to you accessing appropriate related medical information (such as radiological images) for the purposes of this study. ☐

5. I do/do not agree for you to share my anonymised information with external collaborators. ☐

6. I do/do not agree to my hospital number being used to track my data on your secure system. ☐

7. I do/do not agree to my GP being informed of my participation in the study. ☐

8. I agree to take part in the above study. ☐

Name of Patient: _____
(Please print)

Signature: _____ Date: _____
I confirm that I have fully explained the experimental protocol and purpose of the study

Name of Researcher: _____

Signature: _____ Date: _____

Name of person taking consent: _____
(If different from researcher)

Signature: _____ Date: _____

12.1.7 Patient consent form 2

See appendix 4 (Pre methods) “Shoulder ultrasound reproducibility study patient consent form”

12.2 Patient data collection form

Section 1:

Patient code	
Date of 1st appointment	
Treating clinician name	
Site:	UHW = 1 W'church = 2
Gender	1 = Male 2 = Female
Age	
Referral source	GP=1 Rheum=2 Ortho=3 Other = 4
If referral source is "Other", then who?	

Section 2:
Demographics

Currently working? ***Circle all that apply***	1 = full time work 2 = part time work, 3 = retired 4 = full time carer 5 = part time carer 6 = student or unemployed
What is their paid job name	
Occupational factors: How often carry 10 kilos at work?	1 = seldom/never 2 = sometimes 3 = extremely/often 888=Not applicable (as not working)
Occupational factors: How often work above shoulder level?	1 = seldom/never 2 = sometimes 3 = extremely/often 888=Not applicable (as not working)
Education	1 = College/University, 2 = ≤ 12 years in school
Sports – overhead? (e.g. tennis, swimming, basketball, etc)	1=Yes 2=No
Type of overhead Sport?	
Level of overhead Sport?	1 = doesn't play 2 = recreational 3 = club 4 = elite
Additional recreational activity that involves overhead activity?	1=Yes 2=No
If so, what? ***Free hand description***	

Arm side

Handed	1=L 2=R 3=Ambidextrous
Side of symptoms(Sx)	1=L 2=R 3=Bilateral
If bilateral, which is the side they have been referred for / worst side?	1=L 2=R 3= L&R equally 888=not applicable
Involvement of the dominant side?	1=Yes 2=No

Presenting Condition

Duration of current episode at presentation to Physiotherapy	1 = 0-3 months 2 = 3-6 months 3 = 6-12 months 4 = 12-24 months 5 = >24 months
Previous shoulder pain (prior to this episode?)	1=Yes 2=No
Precipitating cause	1 = Unknown / insidious / gradual onset 2 = Injury / trauma 3 = Strain, overuse: unusual activities 4 = Strain, overuse: usual activities
Precipitating cause ***Free hand description***	
Course: Is it getting……..	1= Worse 2= Same 3= Better
Associated neck pain?	1=Yes 2=No
Neurological symptoms?	1=None 2 = Pins and needles / Tingling 3 = Distal symptoms

Investigations (previous or booked):

Type	1=None, 2 = Ultrasound, 3 = MRI, 4= X-ray, 5= Other (if so what?)
Details / findings	

Previous treatment:

Treatment via GP	1 = no contact 2 = advice only, 3 = oral medication and/or topical medication (Analg / NSAIDs) (+/- advice) 4 = oral medication + injection 5 = injection only 6=referral only
Number of injections via GP	
Injection by non-GP	1=Yes 2=No
Previous rehab (specifically for the shoulder)?	0=None, 1=Physio, 2 = Chiro, 3 = Osteo, 4 = Combination
What type / frequency of rehab and how successful?	

Previous Medical History

Diabetes mellitus	1=Yes 2=No
Other medical condition? (e.g. Cancer, heart disease, Osteoarthritis, Rheumatoid arthritis, etc) If so, what?	1=Yes 2=No
Smoker?	1=Current 2=Previous 3=Never

DH:

Currently taking pain killers / NSAIDs on "regular" basis (specifically for the shoulder pain they have been referred to Physiotherapy for)?	1=Yes 2=No

Section 3

Anthropometrics

Height (m):	Weight:	BMI (kg/m2):

Circle painful side

ROM

	Left	**Right**
End of range (EoR)		
Onset of Sx:		
Painful during movement:		

Scapular videoing Weighted performed? 1=Yes 2=No	Weight = (Anything of note……)

Strength (Anything of note)

Distance elbow to radial styloid (m):	Distance GHJ to radial styloid (m):

Circle painful side

	Left	Mean	**Right**	Mean
IR				
ER				
Scaption				

Ortho

Hawkins-Kennedy	Neer
1= +ve 2= -ve	1= +ve 2= -ve

Empty can / resisted isometric abduction/scaption	Painful arc	Pain with active sh elevation
1= +ve 2= -ve	1= +ve 2= -ve	1= +ve 2= -ve

Shoulder diagnostic ultrasound scan patient findings

Side: *L* *R*

Biceps Tendon (Steps 1,2,3):

1=Normal *2=Tendinosis* *3=Tenosynovitis* *4=Tear* *5=Displaced / dislocating*

Subscapularis (Steps 4,5):

1=Normal *2=Tendinosis* *3= Partial tear* *4 = Complete tear*

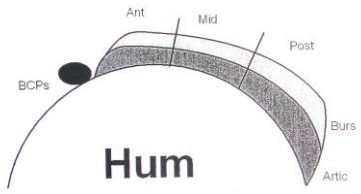
ACJ (Step 6):

1=Normal *Osteophytic 2=mild / 3=mod / 4=severe* *5=Tenosynovitis / Inflammation*

Supraspinatus (Steps 7,8):

1=Normal *2=Tendinosis* *3= Partial tear* *4 = Complete tear*

Bursal thickening / increased fluid = *yes* *no*



Infraspinatus (Steps 9,10):

1=Normal *2=Tendinosis* *3= Partial tear* *4 = Complete tear*

Overall: Tear presence / type

Rotator cuff tear present: 1=**Yes** **2=No**

If present: 1=**Partial thickness tear (PTT)** **2=Full thickness tear (FTT)**

12.3 Order of testing

The series of questionnaires which were posted to patients at the same time as the invitation to be involved with the study are listed below.

- C2. Patient reported measures: Psychological symptoms (4DSQ)
- C3. Patient reported measures: Function / Disability (SPADI)
- Outcome measure (OSS)

Those patients who decided to participate in the study were instructed only to complete the questionnaires in the pack immediately prior to their first appointment.

<u>On day of data collection:</u>

1. Meet and greet patient; answer any questions regarding ethics and consent
2. Complete ethics form 1 and form 2
3. Check questionnaire pack: answer any questions, assist patient in completing any missing sections
Where elevated 4DSQ scores recommend that patient makes follow up contact with their GP
4. Complete VAS

Physical measures:
5. ROM
6. Scapular dyskinesis +/- weighted
7. Strength
8. Orthopaedics tests
9. Ultrasound scan

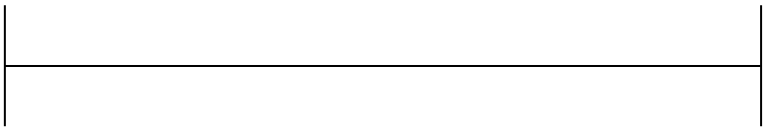
12.4 Visual Analogue Scale

Pain is a complex aspect of human life, and measuring it is challenging. One way for us to quantify your pain is for you to represent your pain on the diagram below.

The horizontal line is used as a sliding scale of your pain. The left hand side of the scale represents no pain and the right hand side represents the worst pain imaginable.

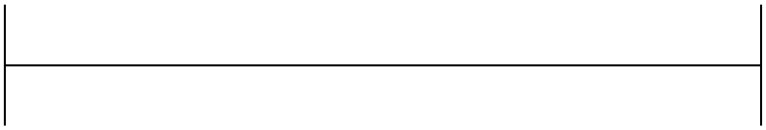
We would like you to consider your pain and place a single mark along the scale at a point you feel represents your pain:

On activity:
(for example *your* painful movement)



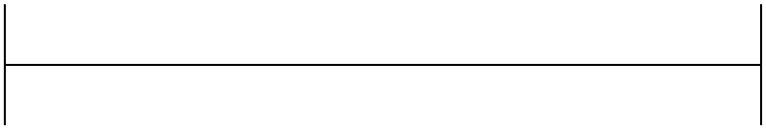
No Pain — Worst Pain imaginable

At rest:
(for example when you are just resting with your arm by your side)



No Pain — Worst Pain imaginable

At night over the last week:



No Pain — Worst Pain imaginable

12.5 4DSQ: Patient to complete

The following is a list of questions about various complaints and symptoms you may have. Each question refers to the complaints and symptoms that you had **in the past week (the past 7 days, including today)**. Complaints you had before then, but no longer had during the past week, do not count.
Please indicate for each complaint how often you noticed that you had it in the past week by putting an "X" in the box under the answer that is most appropriate.

	no	sometimes	regularly	often	very often or constantly
During <u>the past week</u>, did you suffer from:					
1. dizziness or feeling light-headed?	☐	☐	☐	☐	☐
2. painful muscles?	☐	☐	☐	☐	☐
3. fainting?	☐	☐	☐	☐	☐
4. neck pain?	☐	☐	☐	☐	☐
5. back pain?	☐	☐	☐	☐	☐
6. excessive sweating?	☐	☐	☐	☐	☐
7. palpitations?	☐	☐	☐	☐	☐
8. headache?	☐	☐	☐	☐	☐
9. a bloated feeling in the abdomen?	☐	☐	☐	☐	☐
10. blurred vision or spots in front of your eyes?	☐	☐	☐	☐	☐
11. shortness of breath?	☐	☐	☐	☐	☐
12. nausea or an upset stomach?	☐	☐	☐	☐	☐
During <u>the past week</u>, did you suffer from:					
13. pain in the abdomen or stomach area?	☐	☐	☐	☐	☐
14. tingling in the fingers?	☐	☐	☐	☐	☐
15. pressure or a tight feeling in the chest?	☐	☐	☐	☐	☐
16. pain in the chest?	☐	☐	☐	☐	☐
17. feeling down or depressed?	☐	☐	☐	☐	☐
18. sudden fright for no reason?	☐	☐	☐	☐	☐
19. worry?	☐	☐	☐	☐	☐
20. disturbed sleep?	☐	☐	☐	☐	☐
21. a vague feeling of fear?	☐	☐	☐	☐	☐
22. lack of energy?	☐	☐	☐	☐	☐
23. trembling when with other people?	☐	☐	☐	☐	☐
24. anxiety or panic attacks?	☐	☐	☐	☐	☐
During <u>the past week</u>, did you feel:					
25. tense?	☐	☐	☐	☐	☐
26. easily irritated?	☐	☐	☐	☐	☐
27. frightened?	☐	☐	☐	☐	☐

	no	sometimes	regularly	often	very often or constantly
During the past week, did you feel:					
28. that everything is meaningless?	☐	☐	☐	☐	☐
29. that you just can't do anything anymore?	☐	☐	☐	☐	☐
30. that life is not worth while?	☐	☐	☐	☐	☐
31. that you can no longer take any interest in the people and things around you?	☐	☐	☐	☐	☐
32. that you can't cope anymore?	☐	☐	☐	☐	☐
33. that you would be better off if you were dead?	☐	☐	☐	☐	☐
34. that you can't enjoy anything anymore?	☐	☐	☐	☐	☐
35. that there is no escape from your situation?	☐	☐	☐	☐	☐
36. that you can't face it anymore?	☐	☐	☐	☐	☐
During the past week, did you:					
37. no longer feel like doing anything?	☐	☐	☐	☐	☐
38. have difficulty in thinking clearly?	☐	☐	☐	☐	☐
39. have difficulty in getting to sleep?	☐	☐	☐	☐	☐
40. have any fear of going out of the house alone?	☐	☐	☐	☐	☐
During the past week:					
41. did you easily become emotional?	☐	☐	☐	☐	☐
42. were you afraid of anything when there was really no need for you to be afraid? *(for instance animals, heights, small rooms)*	☐	☐	☐	☐	☐
43. were you afraid to travel on buses, streetcars/ trams, subways or trains?	☐	☐	☐	☐	☐
44. were you afraid of becoming embarrassed when with other people?	☐	☐	☐	☐	☐
45. did you ever feel as if you were being threatened by unknown danger?	☐	☐	☐	☐	☐
46. did you ever think "I wish I was dead"?	☐	☐	☐	☐	☐
47. did you ever have fleeting images of any upsetting event(s) that you have experienced?	☐	☐	☐	☐	☐
48. did you ever have to do your best to put aside thoughts about any upsetting event(s)?	☐	☐	☐	☐	☐
49. did you have to avoid certain places because they frightened you?	☐	☐	☐	☐	☐
50. did you have to repeat some actions a number of times before you could do something else?	☐	☐	☐	☐	☐

12.6 4DSQ: Scoring and interpretation

4DSQ scoring form

'no' = score 0
'sometimes' = score 1
'regularly' or more often = score 2

Distress		Depression		Anxiety		Somatisation	
17		28		18		1	
19		30		21		2	
20		33		23		3	
22		34		24		4	
25		35		27		5	
26		46		40		6	
29				42		7	
31				43		8	
32				44		9	
36				45		10	
37				49		11	
38				50		12	
39						13	
41						14	
47						15	
48						16	

Interpretation:

	Distress	Depression	Anxiety	Somatisation
moderately elevated:	> 10	> 2	> 3	> 10
strongly elevated:	> 20	> 5	> 8*	> 20

12.7 SPADI

Shoulder Pain and Disability Index (SPADI)

Source: Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. *Arthritis Care Res*. 1991 Dec;4(4):143-9.

The Shoulder Pain and Disability Index (SPADI) is a self-administered questionnaire that consists of two dimensions, one for pain and the other for functional activities. The pain dimension consists of five questions regarding the severity of an individual's pain. Functional activities are assessed with eight questions designed to measure the degree of difficulty an individual has with various activities of daily living that require upper-extremity use. The SPADI takes 5 to 10 minutes for a patient to complete and is the only reliable and valid region-specific measure for the shoulder.

Scoring instructions

To answer the questions, patients place a mark on a 10cm visual analogue scale for each question. Verbal anchors for the pain dimension are 'no pain at all' and 'worst pain imaginable', and those for the functional activities are 'no difficulty' and 'so difficult it required help'. The scores from both dimensions are averaged to derive a total score.

Interpretation of scores

Total pain score: _____/ 50 x 100 = %

(Note: If a person does not answer all questions divide by the total possible score, eg. if 1 question missed divide by 40)

Total disability score: _____/ 80 x 100 = %

(Note: If a person does not answer all questions divide by the total possible score, eg. if 1 question missed divide by 70)

Total Spadi score: _____/ 130 x 100 = %

(Note: If a person does not answer all questions divide by the total possible score, eg. if 1 question missed divide by 120)

The means of the two subscales are averaged to produce a total score ranging from 0 (best) to 100 (worst).

Minimum Detectable Change (90% confidence) = 13 points

(Change less than this may be attributable to measurement error)

Shoulder Pain and Disability Index (SPADI)

Please place a mark on the line that best represents your experience during the last week attributable to your shoulder problem.

Pain scale

How severe is your pain?

Circle the number that best describes your pain where: 0 = no pain and 10 = the worst pain imaginable.

At its worst?	0	1	2	3	4	5	6	7	8	9	10
When lying on the involved side?	0	1	2	3	4	5	6	7	8	9	10
Reaching for something on a high shelf?	0	1	2	3	4	5	6	7	8	9	10
Touching the back of your neck?	0	1	2	3	4	5	6	7	8	9	10
Pushing with the involved arm?	0	1	2	3	4	5	6	7	8	9	10

Disability scale

How much difficulty do you have?

Circle the number that best describes your experience where: 0 = no difficulty and 10 = so difficult it requires help.

Washing your hair?	0	1	2	3	4	5	6	7	8	9	10
Washing your back?	0	1	2	3	4	5	6	7	8	9	10
Putting on an undershirt or jumper?	0	1	2	3	4	5	6	7	8	9	10
Putting on a shirt that buttons down the front?	0	1	2	3	4	5	6	7	8	9	10
Putting on your pants?	0	1	2	3	4	5	6	7	8	9	10
Placing an object on a high shelf?	0	1	2	3	4	5	6	7	8	9	10
Carrying a heavy object of 10 pounds (4.5 kilograms)	0	1	2	3	4	5	6	7	8	9	10
Removing something from your back pocket?	0	1	2	3	4	5	6	7	8	9	10

12.8 Strength testing

Figure 12-1 IR testing position

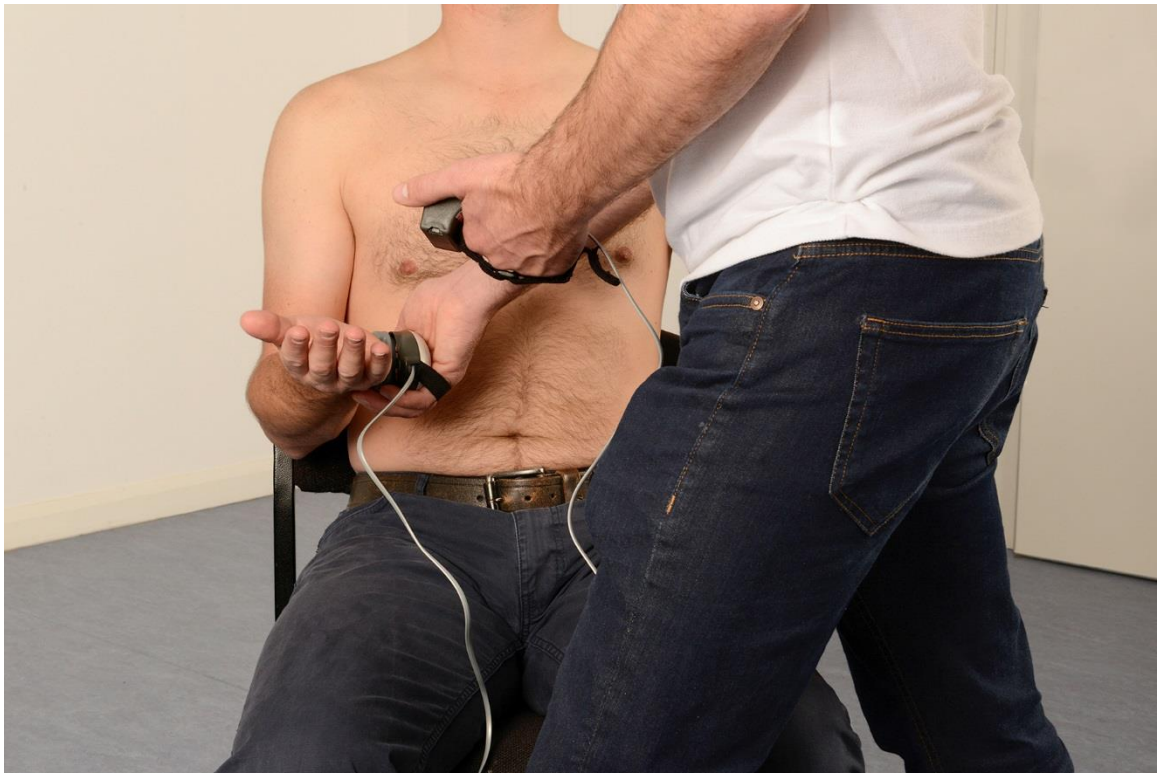
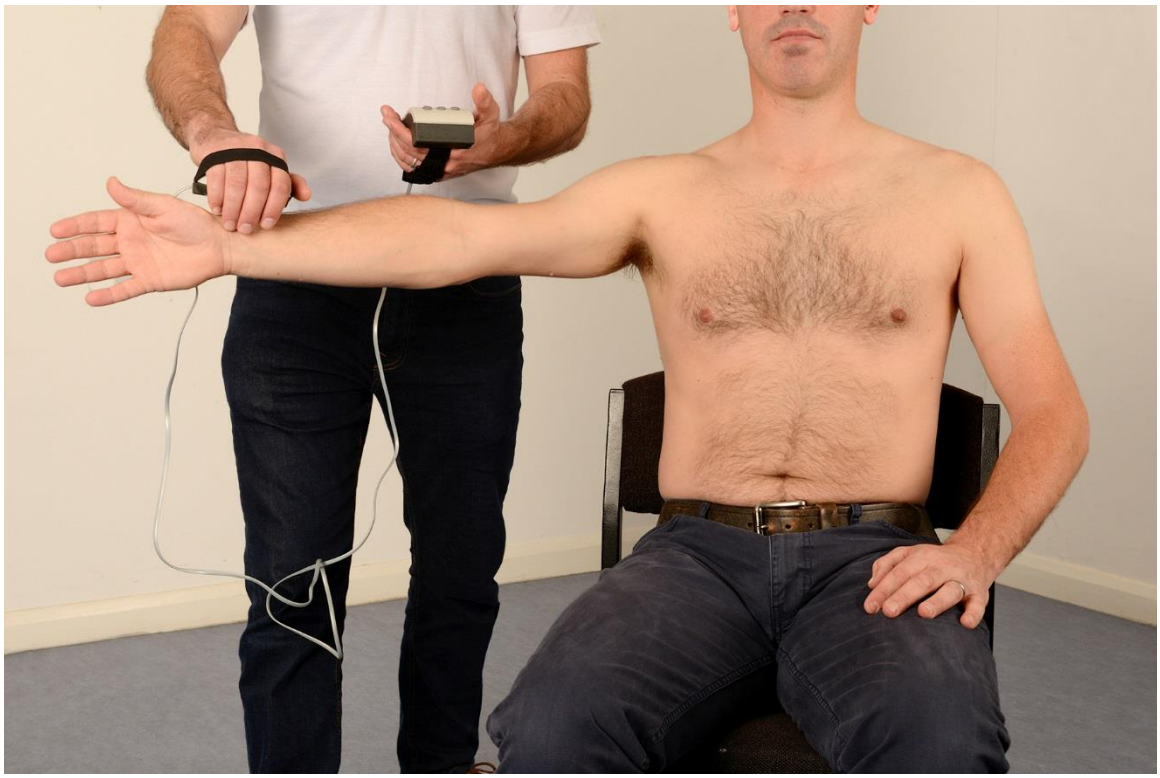


Figure 12-2 Scaption testing position



12.9 ROM testing

Figure* 12-3 *ROM testing position

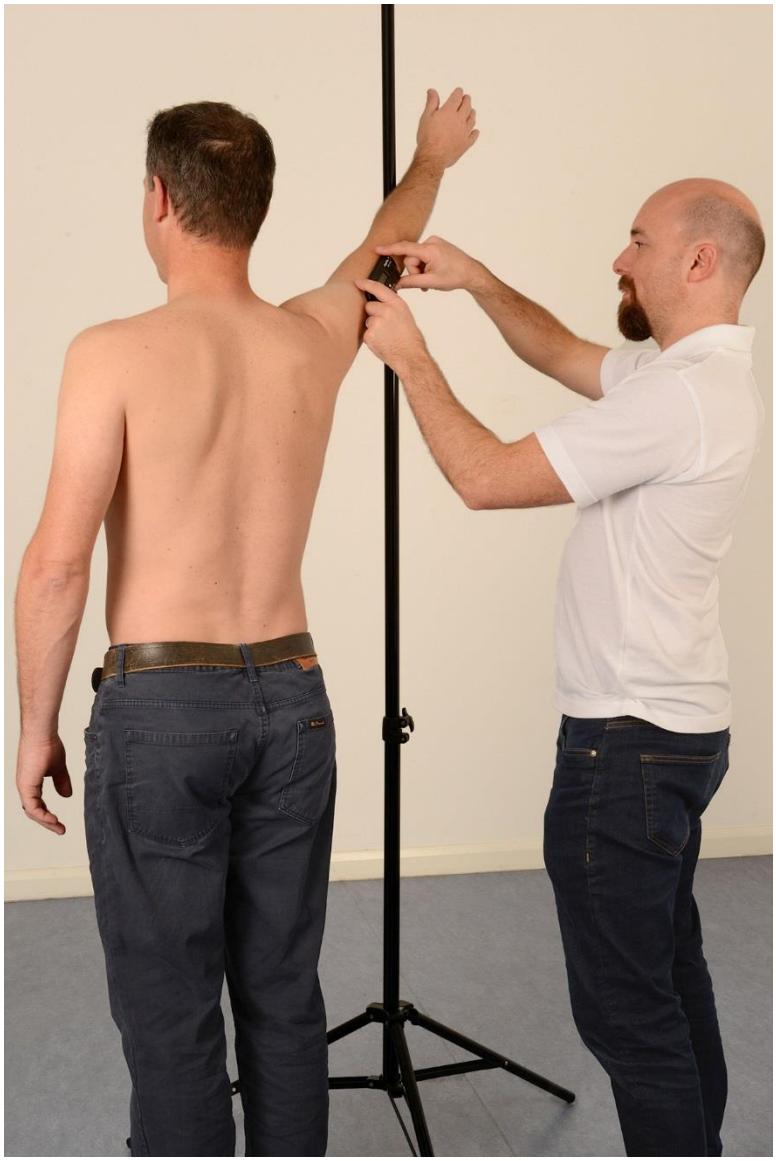


Figure 12-4 ROM testing

12.10 Scapular Dyskinesis test assessment criteria

Operational definitions
Normal scapulohumeral rhythm: The scapula is stable with minimal motion during the initial 30° to 60° of humerothoracic elevation, then smoothly and continuously rotates upward during elevation and smoothly and continuously rotates downward during humeral lowering. No evidence of winging is present.
Scapular dyskinesis: Either or both of the following motion abnormalities may be present.

- *Dysrhythmia*: The scapula demonstrates premature or excessive elevation or protraction, nonsmooth or stuttering motion during arm elevation or lowering, or rapid downward rotation during arm lowering.
- *Winging*: The medial border and/or inferior angle of the scapula are posteriorly displaced away from the posterior thorax.

Rating Scale
Each test movement (flexion and abduction) rated as
a) *Normal motion*: no evidence of abnormality
b) *Subtle abnormality*: mild or questionable evidence of abnormality, not consistently present
c) *Obvious abnormality*: striking, clearly apparent abnormality, evident on at least 3/5 trials (dysrhythmias or winging of 1 in [2.54 cm] or greater displacement of scapula from thorax)

Final rating
This is based on combined flexion and abduction test movements:
Normal: Both test motions are rated as normal or 1 motion is rated as normal and the other as having subtle abnormality.
Subtle abnormality: Both flexion and abduction are rated as having subtle abnormalities.
Obvious abnormality: Either flexion or abduction is rated as having obvious abnormality.

McClure et al. (2009)

12.11 Specific criteria relating to the sonographic appearance of cuff tendinopathy and bursitis.

Components of tendinopathy and their grading

Normal tendon	⟺	Tendinosed tendon
"Normal" **thickness** • Caveat = relative to the size of the patient / the activity levels of their shoulder / their uninvolved shoulder	⟺	"Abnormal" **thickness** = grossly thickened / thinned • Caveat = relative to the size of the patient / the activity levels of their shoulder / their uninvolved shoulder
"Normal" **echogenicity** = hyperechoic (bright) • Caveat = acknowledging how easily the patient's tissues do or don't image; also when accommodating anisotropy	⟺	"Abnormal" **echogenicity** = grossly hypoechoic (dark) • Caveat = acknowledging how easily the patient's tissues do or don't image; also when accommodating anisotropy
"Normal" **echotexture** = homogenous with an ordered fibre orientation • Caveat = acknowledging how easily the patient's tissues do or don't image; also when accommodating anisotropy	⟺	"Abnormal" **echotexture** = grossly heterogenous; calcific deposits • Caveat = acknowledging how easily the patient's tissues do or don't image; also when accommodating anisotropy

Components of bursal involvement and their grading

Normal bursa	⟺	Abnormal bursa
"Normal" **thickness** • Less than 2mm thick	⟺	"Abnormal" **thickness** = grossly thickened
"Normal" **fluid** • Less than 2mm of hypoechoic fluid in the bursa	⟺	"Abnormal" **fluid** = grossly increased "bursal effusion"

12.12 Orthopaedic tests

Figure 12-5 Neer's sign

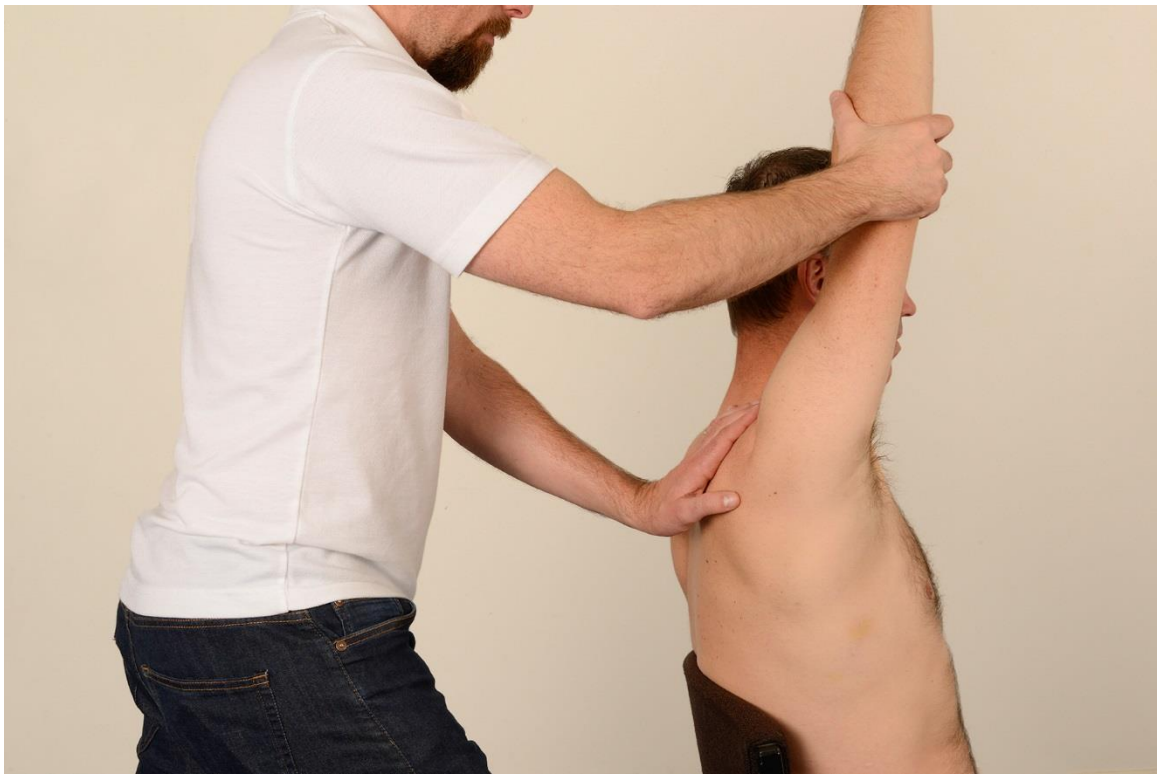


Figure 12-6 Empty can test

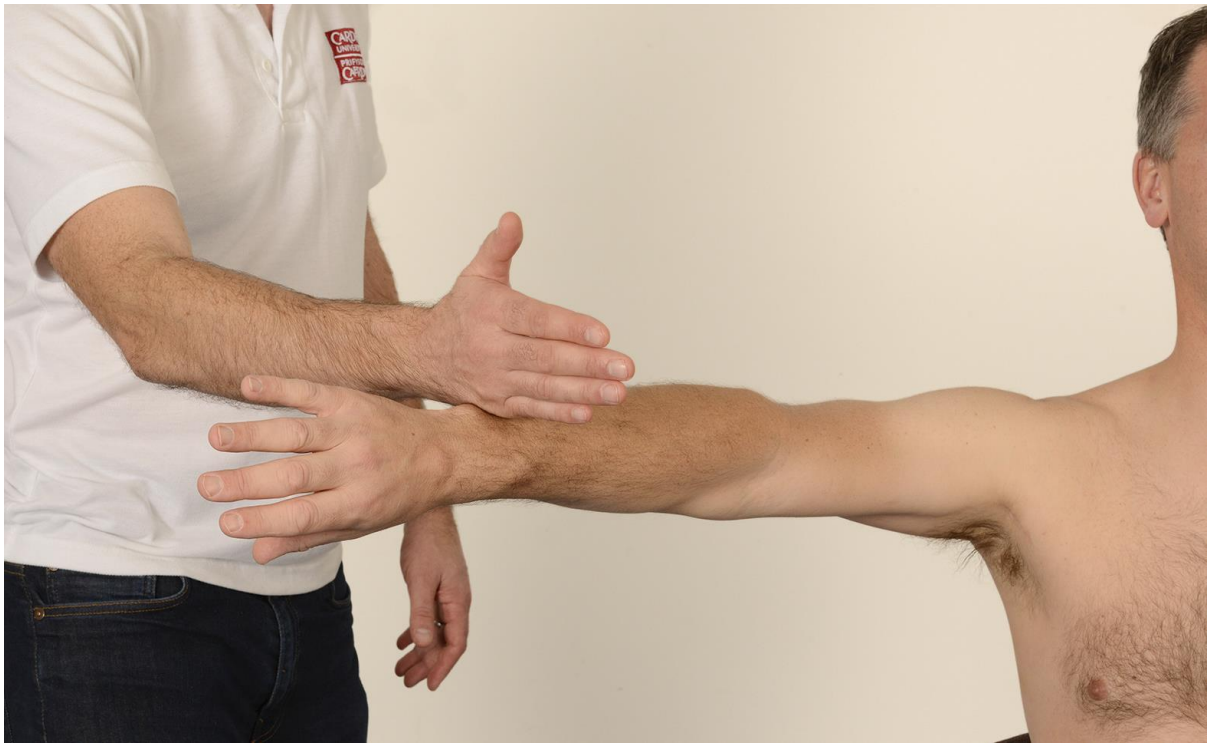
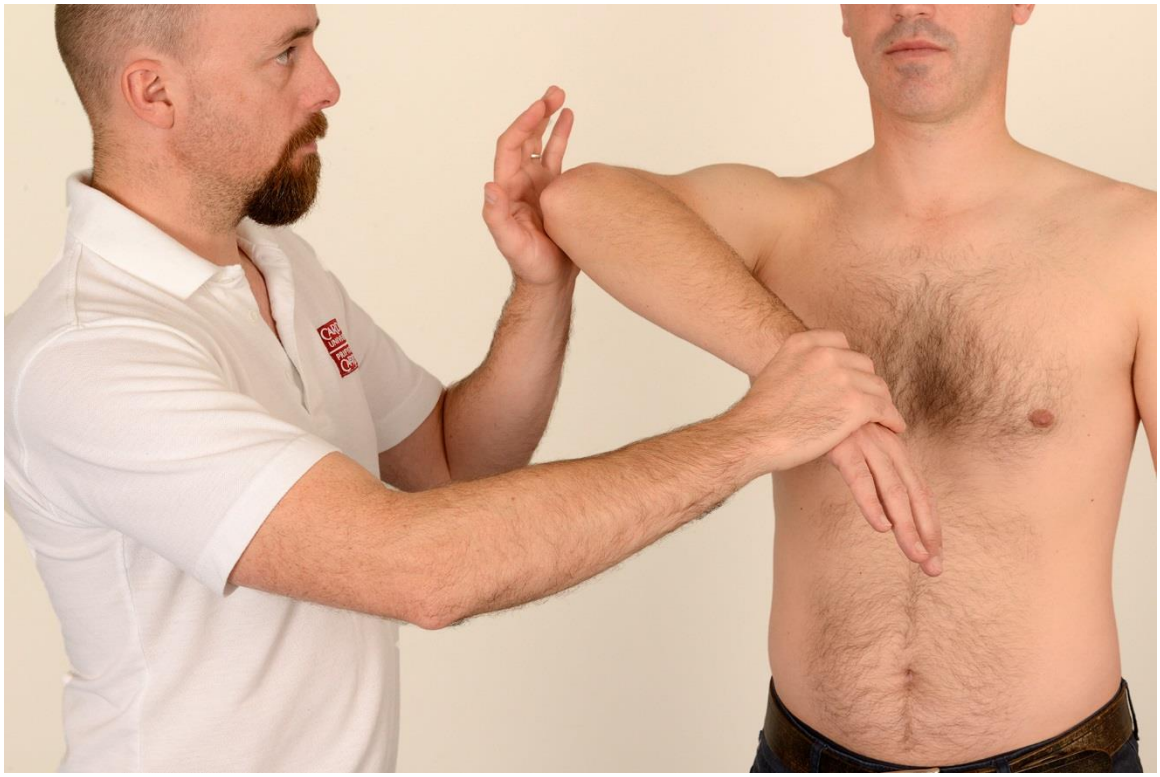


Figure 12-7 Hawkins-Kennedy test

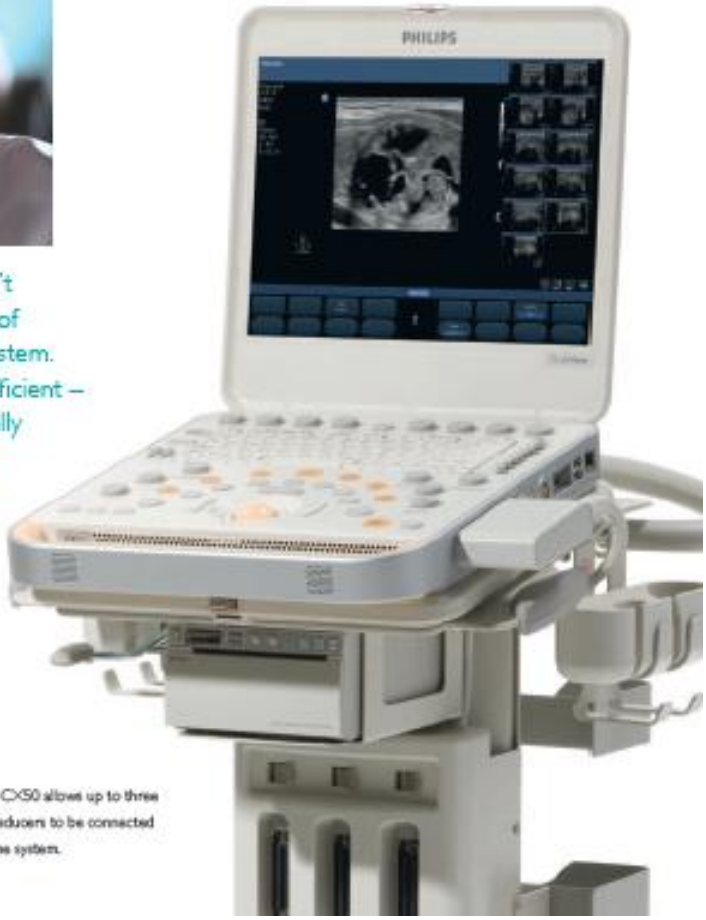


12.13 Philips CX-50

Philips CX50 CompactXtreme system

Now you can be confident in the data from your exams, including your most technically challenging studies. Philips, a leader in cutting-edge ultrasound development, has integrated premium, innovative technologies into the new CX50 system to address the need for premium class performance in a compact ultrasound system.

"While doing portable studies I can't always count on the performance of my smaller, compact ultrasound system. The image quality is not always sufficient – especially on difficult patients. I really don't like sacrificing performance for portability. I need a very small, light-weight ultrasound system that provides me the same image quality as my larger ultrasound systems."



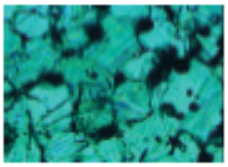
The CX50 allows up to three transducers to be connected to the system.

Premium class performance in a compact system

Extreme performance is built into the CX50 system with clinically proven premium technologies. Images are displayed with exceptional clarity, giving you the data you need for confident diagnoses.

PureWave goes portable

The CX50 CompactXtreme ultrasound system features PureWave crystal technology, one of the biggest breakthroughs in piezoelectric transducer technology in 40 years. With PureWave crystal technology, clinicians can rely on exceptional tissue detail, enhanced far field resolution, and the ability to image a wide variety of patients, including the technically difficult.



Conventional

PureWave

As you can see (800X magnification), PureWave crystals have virtually perfect uniformity for greater bandwidth and twice the efficiency of conventional ceramic materials. The result is excellent imaging and Doppler performance.

Transducers for a full range of clinical applications

The CX50 supports premium performance 2D and Doppler for a wide range of exam needs – abdominal, vascular, breast, MSK, pediatrics, superficial, obstetrical, gynecological, surgical, and cardiac.

- C5-1 PureWave curved array
- C10-3v PureWave curved array
- C8-5 broadband curved array
- C9-3io broadband curved array
- L12-5/50 broadband linear array
- L10-4lap broadband linear array
- L12-3 broadband linear array
- L15-7io broadband linear array
- S5-1 PureWave sector array
- S8-3 broadband sector array
- S12-4 broadband sector array
- X7-2t PureWave TEE

Digital broadband beamforming on a compact

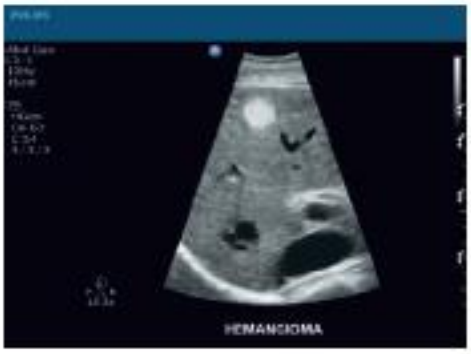
The CX50 combines the broadband capabilities of a digital beamformer with the broadband signals produced by Philips premium transducers including those with PureWave technology. For the first time in a compact system, complete tissue signatures are captured, preserved, and displayed. The level of image quality is exceptional, allowing you to fully appreciate subtle anatomical details.

SonoCT and XRES technologies bring a new level of clarity to compact ultrasound

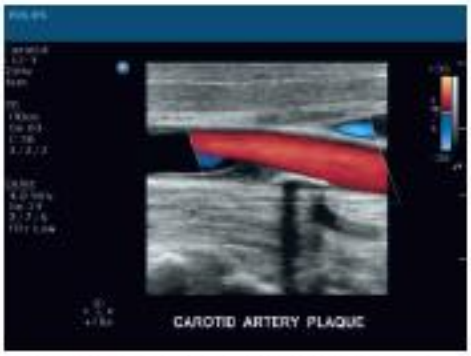
Philips SonoCT is a clinically-proven premium technology that acquires up to nine lines of sight and combines the individual images into one clear well-defined image in real time. SonoCT displays striking levels of tissue differentiation that are virtually free of artifact.

Advanced XRES adaptive image processing reduces speckle, haze, and clutter, resulting in images virtually free from noise with extraordinary clarity and edge definition. When SonoCT and XRES work in tandem, even the subtlest of diagnostic features are enhanced.

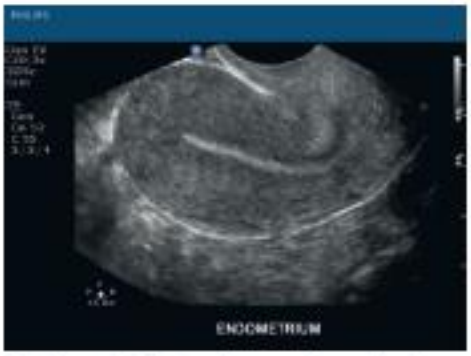
With three surgery transducers on the CX50, you can bring premium performance into the surgery suite.



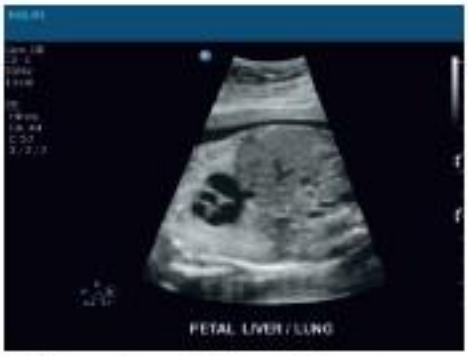
The exquisite detail of this hemangioma and surrounding tissue and structures is the result of PureWave technology on the C5-1 transducer.



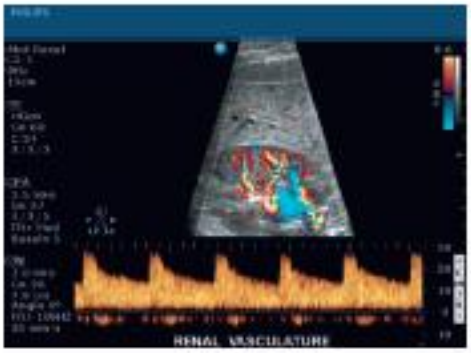
The size and location of plaque and blood flow in the carotid artery are easily appreciated in this image with the L12-3 linear array transducer.



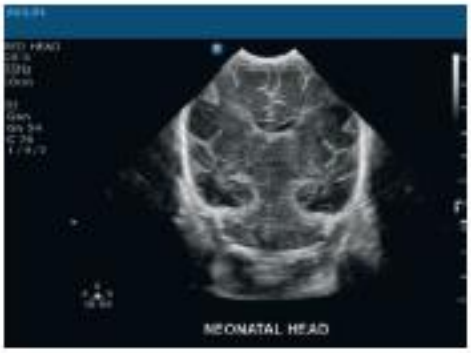
PureWave on the C10-3v transducer provides exceptional clarity of uterine anatomy with clear definition of the midline and walls.



The C5-1 transducer uses PureWave technology to capture details of small structures, such as the fetal anatomy shown in this image of a tiny heart and abdominal organs.



PureWave technology on the C5-1 curved array transducer provides excellent 2D imaging and Doppler sensitivity as seen in this native kidney.



This coronal view of a neonatal head demonstrates the superb clarity provided by the new C8-5 transducer.

Premium compact ultrasound everywhere you need it

The CX50 system adapts to your workflow, whether you're in the ICU, at the bedside, in the ED, or at a remote location. With easy-to-use tools designed for your needs, you're ready to scan wherever your patients are located.

Breakthrough workflow solutions

The CX50 was designed to make portable exams easy and efficient. With a single button, iSCAN technology automatically samples data for quick optimization for 2D and Doppler performance. To increase efficiency and diagnostic confidence, the CX50 incorporates the latest 3D freehand acquisition to enable advanced volume workflow studies.

One-button controls are logically placed on the CX50 control panel for quick selection and optimization during every exam.

Reduce exam time by up 50% with SmartExam

SmartExam protocols are easy-to-use customizable guides that help you perform complete studies on every patient. The on-screen menu guides you through the required views for a specific exam type, automatically enters annotation, and builds your report. Save time, reduce repeated moves, and increase efficiency and consistency of exams.

Compact ultrasound designed for your environment

The CX50 system features a high resolution monitor for optimal viewing in difficult portable environments, and fast system start-up allows you to quickly begin your studies. Wireless and wired DICOM allow flexibility when connecting to your PACS. You can also export your data by DVD and USB media with integrated DICOM viewer.

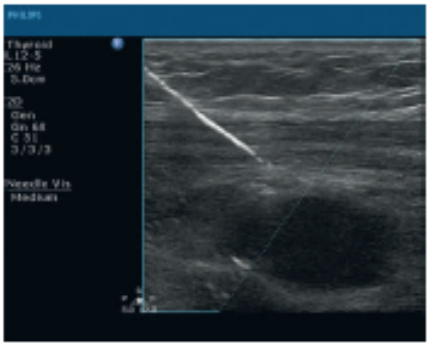
Fine-tune exams with active native data

The CX50 system stores active native acoustic data giving you the ability to adjust virtually all scanning parameters on single images, clips or stored 2D and Doppler data. Images can be readjusted during or after the exam, enhancing diagnostic details, allowing for shorter exam times, and reducing the need for repeat studies.

Expand diagnostic information with QLAB quantification software

The CX50 offers assessment and analysis capabilities with QLAB's clinically proven plug-ins.

- QI 3D quantification – QI 3DQ
- Region of interest – ROI
- Intima media thickness evaluation – IMT
- Cardiac motion quantification with speckle tracking technology – CMQ
- Strain quantification – SQ
- MicroVascular Imaging – MVI



Needle visualization on the L12-5 and L12-3 enhances the presentation of the needle without degrading the surrounding tissue image.

12.14 OSS: Questionnaire

Please tick one box for every question.

1. During the past 4 weeks…

How would you describe the **worst** pain you had from your shoulder?

None	Mild	Moderate	Severe	Unbearable

2. During the past 4 weeks…

Have you had any trouble dressing yourself because of your shoulder?

No trouble at all	A little bit of trouble	Moderate trouble	Extreme difficulty	Impossible to do

3. During the past 4 weeks…

Have you had any trouble getting in and out of a car or using public transport because of your shoulder?

No trouble at all	A little bit of trouble	Moderate trouble	Extreme difficulty	Impossible to do

4. During the past 4 weeks…

Have you been able to use a knife and fork - at the same time?

Yes, easily	With little difficulty	With moderate difficulty	With extreme difficulty	No, impossible

5. During the past 4 weeks…

Could you do the household shopping on your own?

Yes, easily	With little difficulty	With moderate difficulty	With extreme difficulty	No, impossible

6. During the past 4 weeks…

Could you carry a tray containing a plate of food across a room?

Yes, easily	With little difficulty	With moderate difficulty	With extreme difficulty	No, impossible

7. During the past 4 weeks…

Could you brush/comb your hair with the affected arm?

Yes, easily	With little difficulty	With moderate difficulty	With extreme difficulty	No, impossible

8. During the past 4 weeks…

How would you describe the pain you usually had from your shoulder?

None	Very mild	Mild	Moderate	Severe

9. During the past 4 weeks…

Could you hang your clothes up in a wardrobe, using the affected arm?

Yes, easily	With little difficulty	With moderate difficulty	With great difficulty	No, impossible

10. During the past 4 weeks…

Have you been able to wash and dry yourself under both arms?

Yes, easily	With little difficulty	With moderate difficulty	With extreme difficulty	No, impossible

11. During the past 4 weeks…

How much has pain from your shoulder interfered with your usual work (including housework)?

Not at all	A little bit	Moderately	Greatly	Totally

12. During the past 4 weeks…

Have you been troubled by pain from your shoulder in bed at night?

No nights	Only 1 or 2 nights	Some nights	Most nights	Every night

12.15 OSS: Scoring

New scoring system for the Oxford shoulder score*

When the Oxford shoulder score was originally devised, the scoring system was designed to be as simple as possible, in order to encourage its use. Thus, in the original publication (Dawson J, Fitzpatrick R, Carr A. Questionnaire on the Perceptions of Patients About Shoulder Surgery. J Bone Joint Surg [Br] 1996; 78: 593-600) each question was scored from 1 to 5, with 1 representing best outcome/least symptoms. Scores from each question were added so the overall score was from 12 to 60 with 12 being the best outcome. This was the same method as used for the Oxford hip and knee scores - which many surgeons reported finding unintuitive. This led some to adapt the scoring - leading to considerable confusion. We therefore issued recent recommendations concerning changes to the method of scoring the Oxford hip and knee scores[1]. Our view is that similar changes should be implemented with the OSS **and therefore recommend the following**.

Under the new system, each question on the OSS should be scored 0 to 4, with 4 representing the best (this is the opposite direction from the original method of scoring). When the 12 items are summed, this produces overall scores that run from 0 to 48 with 48 being the best outcome (to convert the 'old system' of 60-12 to the 0-48 scoring system and vice versa simply subtract the score from 60)[2]. In addition, the method used should always be clearly stated (including in abstracts). **We also recommend that this scoring system shoulder be adopted for the Oxford Shoulder Instability Score[3].**
To further avoid confusion, always state clearly the method that has been used (including in abstracts).

New system of scoring (more detail)
Each of the 12 questions on the Oxford shoulder score is scored in the same way with the score decreasing as the reported symptoms increase (ie. become worse). All questions are laid out similarly with response categories denoting least (or no) symptoms being to the left of the page (scoring 4) and those representing greatest severity lying on the right hand side (scoring 0). eg. question 1:

1. *During the past 4 weeks.....*
How would you describe the *worst* pain you had from your shoulder?

None	Mild	Moderate	Severe	Unbearable
❑	❑	❑	❑	❑
4	**3**	**2**	**1**	**0**

The overall score is reached by simply summing the scores received for individual questions. This results in a continuous score ranging from 0 (most severe symptoms) to 48 (least symptoms).

Missing values/notes for analysis.
We also propose that, if, after repeated attempts to obtain complete data from an individual, only one or two questions have been left unanswered, it is reasonable to enter the mean value representing all of their other responses, to fill the gaps. If more than two questions are unanswered we recommend that an overall score should not be calculated. If patients indicate two answers for one question we recommend that the convention of using the worst (most severe) response is adopted.

1. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, Dawson J. The Use of the Oxford Hip and Knee Scores. J Bone Joint Surg [Br] 2007; 89-B: 1010-4.
2. Weale AE, Halabi OA, Jones PW, While SH. Perceptions of Outcomes After Unicompartmental and Total Knee Replacements. Clin Orthop Rel Res 2001; 382: 143-53.
3. Dawson J, Fitzpatrick R, Carr A. The Assessment of Shoulder Instability: the Development and Validation of a Questionnaire. J Bone Joint Surg [Br] 1999; 81-B: 420-6.

13 Appendix VI: Results chapter

Prognostic cohort study subjects:

- A. Demographics (raw data)
- B. Clinical history (raw data)

C1. Patient reported measures: Pain

C2. Patient reported measures: 4DSQ

C3. Patient reported measures: Function / disability

D1. Clinical measures: Strength

D2. Clinical measures: ROM

D3. Clinical measures: Scapular movement and control

E1. Structural pathology via imaging

E2. Structural pathology via orthopaedic tests

- Orthopaedic tests: 3 or more +ve tests

Outcome

Treatment

- Raw treatment data
- Number of weeks under treatment and discharge situation

13.1 A. Demographics (raw data)

Table 13-1: Generic demographic variables (n = 76 at baseline)

Subject code	Age (years)	BMI (kgm^{-2})	Gender	Education level	Smoking history
F6847h	51	35.4	2	2	3
F7047h	47	26.3	2	1	3
M2146h	54	37.3	1	2	n/r
F5446h	48	18.2	2	1	n/r
F4417h	50	50.5	2	1	3
M0379h	62	26.4	1	2	2
F4583h	63	26.5	2	1	3
F7849h	50	23.8	2	1	3
M0825h	57	29.5	1	1	3
M0061h	62	31.2	1	1	3
F2520h	47	40.3	2	1	3
F8486h	50	32.2	2	1	3
F1005w	45	35.1	2	1	1
F6416h	68	27.4	2	1	3
F4764h	52	22.1	2	1	3
M4542h	49	29.4	1	1	2
M8878h	57	23.0	1	2	2
F0165h	66	25.6	2	1	2
F5367w	64	35.5	2	1	3
F7405h	38	20.5	2	1	2
M2747w	50	27.7	1	1	3
F1634w	30	42.2	2	1	3
F0738w	47	33.0	2	1	3
M0035w	44	25.5	1	1	3
F5503w	38	40.2	2	1	2
M9819h	29	23.0	1	2	3
M7535w	25	21.4	1	2	1
F0809h	48	32.0	2	2	3
F4113w	35	28.8	2	1	3
F9939h	50	24.3	2	2	1
M5703h	49	30.5	1	1	3
M1518h	64	31.0	1	1	3
F8774w	65	25.9	2	1	2
M8695h	37	27.1	1	1	2
F7304h	50	30.7	2	1	3
F2387h	66	31.7	2	2	2
F8396w	52	25.7	2	1	3
F4674w	63	23.1	2	1	3
M1243h	62	31.3	1	1	3
M2593w	44	30.3	1	1	2
M3288h	63	31.3	1	1	3
M0519h	31	26.7	1	1	3
M2608w	26	26.8	1	1	3
F8966w	64	28.3	2	1	3

F4943h	56	24.1	2	1	3
F5204w	48	29.6	2	1	3
M1662h	55	38.8	1	2	2
M5465w	65	27.4	1	1	2
M8305w	59	29.0	1	1	3
F0732w	49	27.3	2	1	2
F2122w	19	17.3	2	1	3
M2028W	58	25.0	1	1	1
F4142w	64	26.7	2	1	2
F9240h	63	41.6	2	1	3
F2573h	67	37.8	2	1	3
F4735h	23	17.3	2	1	1
F1965w	29	39.9	2	2	3
M6858w	34	33.3	1	1	3
F9830h	52	35.8	2	1	1
M2279h	71	27.3	1	1	2
M1543h	62	26.0	1	1	3
F1188h	38	32.3	2	1	2
M4091h	28	20.2	1	2	3
F5706h	56	22.9	2	1	3
F1669h	18	30.0	2	2	3
F9261h	53	25.6	2	1	2
M7524w	61	31.8	1	2	3
F3795w	51	22.2	2	1	3
M9231w	53	30.3	1	1	2
F5690h	19	24.4	2	1	1
M3609h	40	21.5	1	1	3
F8809w	48	22.6	2	1	3
F6914w	65	31.8	2	1	3
M3960h	61	27.5	1	1	3
M9287h	42	33.1	1	1	3
F2197w	27	26.8	2	1	2

Key: kgm^{-2} = mass in kilogrammes divided by height in meters squared, Gender: 1 = male; 2 = female, Education level: 1 = educated to college or university level; 2 = Less than 12 years in education, Smoking history: 1 = Current; 2 = Previous; 3 = Never, n/r = not recorded

Table 13-2: Work, recreational or functional task related demographic variables (n = 76 at baseline)

Subject code	Currently working	Paid work type	Carry 10Kg at work?	Work above shoulder height?	Play overhead sports?
F6847h	n/r	3	1	1	2
F7047h	2	2	1	1	2
M2146h	6	3	3	3	2
F5446h	2	2	2	3	2
F4417h	1	2	3	3	1
M0379h	2	3	2	2	1
F4583h	3	1	4	4	2
F7849h	3	1	4	4	1
M0825h	1	2	1	1	1
M0061h	3	1	4	4	1
F2520h	1	2	1	1	2
F8486h	1	2	2	2	2
F1005w	1	2	2	2	2
F6416h	3	1	4	4	1
F4764h	2	2	1	1	1
M4542h	1	2	1	1	1
M8878h	1	3	2	3	2
F0165h	3	1	4	4	2
F5367w	2	2	3	3	1
F7405h	1	2	2	1	2
M2747w	1	2	1	1	1
F1634w	2 and 6	1 and 2	1	2	1
F0738w	2 and 6	1 and 2	1	1	1
M0035w	2	2	1	1	1
F5503w	1	2	1	1	2
M9819h	1	3	3	3	1
M7535w	1	3	3	3	2
F0809h	6	n/r	4	4	2
F4113w	1	2	1	1	1
F9939h	1	2	1	2	2
M5703h	1	2	2	2	1
M1518h	1	2	1	1	1
F8774w	6	1	4	4	1
M8695h	1	3	3	3	1
F7304h	2	2	2	2	2
F2387h	3	1	4	4	2
F8396w	1	2	1	1	1
F4674w	2	1	1	2	2
M1243h	3	1	4	4	1
M2593w	1	2	1	1	1
M3288h	1	2	1	1	1
M0519h	6	1	4	4	1
M2608w	1	2	1	2	1

F8966w	3	1	4	4	1
F4943h	2	2	2	1	1
F5204w	1	2	1	1	1
M1662h	6	1	4	4	2
M5465w	1	2	1	1	2
M8305w	1	2	1	1	2
F0732w	1	2	1	1	1
F2122w	6	1	1	1	2
M2028W	6	1	4	4	2
F4142w	3	1	4	4	2
F9240h	4	2	1	1	2
F2573h	3	1	4	4	2
F4735h	1	2	3	3	1
F1965w	6	1	4	4	1
M6858w	1	2	1	1	1
F9830h	1	2	1	1	1
M2279h	3	1	4	4	2
M1543h	2	1	1	1	1
F1188h	1	2	1	1	1
M4091h	1	3	3	1	1
F5706h	2	2	1	1	1
F1669h	4	1	2	1	2
F9261h	1	2	2	2	1
M7524w	2	2	1	1	1
F3795w	1	2	2	2	1
M9231w	1	2	1	1	1
F5690h	6	1	1	2	1
M3609h	1	2	1	1	1
F8809w	2	2	1	1	1
F6914w	3	1	4	4	1
M3960h	3	1	4	4	2
M9287h	1	2	1	1	1
F2197w	1	2	2	2	1

Key: Currently working: Full time = 1; Part time = 2; Retired = 3; Full time carer = 4; Part time carer = 5; Student or unemployed = 6, Paid work type: 1 = Not paid working; 2 = Professional / managerial / office; 3 = Manual / semi-manual / unskilled, How often carry 10Kg at work?: 1 = Seldom / never; 2 = Sometimes; 3 = Extremely / often; 4 = Not applicable, How often work above shoulder height?: 1 = Seldom / never; 2 = Sometimes; 3 = Extremely / often; 4 = Not applicable, Play overhead sports?: 1 = Yes, 2 = No, n/r = not recorded

13.2 B. Clinical history (raw data)

Table 13-3: Current condition related variables (n = 76 at baseline)

Subject code	Duration current episode	Involvement of dominant side	Precipitating cause	Previous shoulder pain	Associated neck pain	Neurological symptoms	Course of condition
F6847h	1	2	1	2	2	1	2
F7047h	4	1	2	2	2	1	2
M2146h	5	1	1	1	2	1	2
F5446h	2	1	2	1	2	2	1
F4417h	2	2	2	2	2	1	1
M0379h	2	2	1	2	n/r	n/r	1
F4583h	3	1	n/r	1	1	2	3
F7849h	2	1	2	2	2	1	3
M0825h	2	2	1	2	2	1	3
M0061h	2	1	1	2	1	n/r	1
F2520h	2	2	1	2	1	1	2
F8486h	5	1	3	2	2	1	1
F1005w	2	2	1	1	2	1	1
F6416h	4	1	1	2	1	1	1
F4764h	3	2	1	2	1	1	1
M4542h	1	1	3	2	2	1	1
M8878h	3	2	1	2	2	3	2
F0165h	2	1	1	1	1	2	3
F5367w	2	2	2	2	1	1	3
F7405h	1	2	1	1	2	2	2
M2747w	5	1	4	2	1	1	2
F1634w	3	2	2	2	1	3	2
F0738w	3	1	2	2	2	1	2
M0035w	3	2	1	1	1	3	2
F5503w	3	2	1	2	2	3	2
M9819h	2	1	4	2	2	1	2
M7535w	3	1	4	2	2	1	1
F0809h	2	2	1	2	2	1	3
F4113w	4	1	1	2	1	2	2
F9939h	3	2	1	2	2	2	2
M5703h	3	2	1	2	2	1	2
M1518h	2	2	1	2	1	3	2
F8774w	3	2	1	2	2	1	2
M8695h	1	1	4	1	2	2	1
F7304h	4	1	1	2	1	1	2
F2387h	2	2	1	2	1	2	3
F8396w	3	1	1	2	1	2	1
F4674w	3	2	1	2	2	1	3
M1243h	2	2	1	2	2	1	2
M2593w	4	1	3	2	2	1	3
M3288h	2	1	1	2	2	1	n/r
M0519h	4	2	3	2	2	1	1
M2608w	3	1	1	2	2	1	2
F8966w	2	1	2	2	2	3	2
F4943h	2	1	2 and 3	1	2	2	3
F5204w	3	2	1	2	2	1	3
M1662h	2	1	1	2	2	3	1

M5465w	2	2	1	1	2	1	1
M8305w	3	2	1	2	1	1	2
F0732w	3	2	1	2	2	1	2
F2122w	5	2	1	2	2	1	2
M2028W	1	2	1	2	2	4	2
F4142w	5	2	1	2	2	n/r	2
F9240h	3	2	1	2	2	3	1
F2573h	5	1	1	2	2	1	2
F4735h	2	2	1	2	2	3	3
F1965w	5	1	2	2	1	2	2
M6858w	4	2	4	2	2	1	2
F9830h	3	2	3	2	2	1	2
M2279h	1	2	1	1	2	1	2
M1543h	3	2	3	2	2	1	2
F1188h	3	2	1	2	2	1	2
M4091h	5	Ambidextrous	2	2	2	1	2
F5706h	5	2	4	2	2	2	2
F1669h	5	1	2	2	2	1	2
F9261h	2	2	1	2	2	1	1
M7524w	5	1	4	2	2	1	2
F3795w	3	1	1	1	2	1	2
M9231w	3	1	1	1	1	1	1
F5690h	3	1	1	2	1	2	3
M3609h	5	1	4	2	2	1	2
F8809w	3	1	1	2	2	1	1
F6914w	2	1	1	1	1	1	1
M3960h	2	2	1	2	2	1	2
M9287h	3	1	4	2	1	2	3
F2197w	4	2	1	1	2	2	1

Key: Duration current episode?: 1 = 0-3 months; 2 = 3-6 months; 3 = 6-12 months; 4 = 12-24 months; 5 = More than 24 months, Involvement of dominant side?: 1 = Yes, 2 = No, Precipitating cause: 1 = Unknown / insidious / gradual onset; 2 = Injury / trauma; 3 = Strain / overuse: unusual activities; 4 = Strain / overuse: usual activities, Previous shoulder pain: 1 = Yes, 2 = No, Associated neck pain?: 1 = Yes, 2 = No, Neurological symptoms: 1 = None; 2 = Pins and needles / tingling; 3 = Distal symptoms; 4 = other, Course of condition: 1 = Worse; 2 = Same; 3 = Better, n/r = not recorded

Table 13-4: Treatment related variables (n = 76 at baseline)

Subject code	Previous rehabilitation	This episode; currently taking analgesia/NSAIDS	This episode; treatment via GP	This episode; number of GP injections	This episode; injection by non GP	This episode; investigations
F6847h	0	1	2	0	2	2
F7047h	0	2	3	0	2	6
M2146h	1	1	2	0	2	6
F5446h	0	2	3	0	2	2
F4417h	1	1	5	1	2	2
M0379h	0	2	3	0	2	1
F4583h	0	1	4	1	2	2
F7849h	0	2	1	0	2	4
M0825h	0	2	6	0	2	1
M0061h	0	2	3	0	2	2
F2520h	0	1	3	0	2	1
F8486h	1	1	5	2	1	6
F1005w	3	1	4	1	2	1
F6416h	1	1	6	0	2	1
F4764h	0	2	3	0	2	4
M4542h	0	1	6	0	2	1
M8878h	0	2	3	0	2	4
F0165h	1	1	5	1	2	1
F5367w	1	2	6	0	2	2
F7405h	0	2	3	0	2	4
M2747w	0	2	3	0	2	2
F1634w	1	1	4	1	2	1
F0738w	0	1	3	0	2	1
M0035w	0	2	6	0	2	1
F5503w	0	2	6	0	2	1
M9819h	3	1	6	0	2	1
M7535w	0	2	6	0	2	1
F0809h	0	2	6	0	2	n/r
F4113w	1	n/r	3	0	2	4
F9939h	0	2	3	0	2	1
M5703h	0	2	3	0	2	1
M1518h	0	2	5	1	2	4
F8774w	0	2	3	0	2	1
M8695h	1	2	6	0	2	1
F7304h	0	2	3	0	2	4
F2387h	0	2	6	0	2	1
F8396w	0	2	1	0	2	1
F4674w	0	2	6	0	2	4
M1243h	0	1	6	0	2	1
M2593w	2	2	6	0	2	2
M3288h	0	2	6	0	2	1
M0519h	0	2	6	0	2	1

M2608w	0	2	6	0	2	1
F8966w	0	1	6	0	2	4
F4943h	2	2	6	0	2	4
F5204w	0	2	3	0	2	1
M1662h	0	1	4	2	2	4
M5465w	0	2	6	0	2	4
M8305w	0	2	6	0	2	1
F0732w	0	1	4	2	2	6
F2122w	0	2	6	0	2	1
M2028W	1	2	3	0	2	1
F4142w	0	2	3	0	2	1
F9240h	1	1	5	1	2	2, 3 and 4
F2573h	1	2	3	0	2	5
F4735h	0	2	3	0	2	4
F1965w	0	2	6	0	2	3 and 4
M6858w	0	2	3	0	2	1
F9830h	0	2	3	0	2	1
M2279h	0	1	4	1	2	1
M1543h	0	2	6	0	2	4
F1188h	0	2	3	0	2	1
M4091h	0	2	3	0	2	1
F5706h	0	2	5	1	2	2
F1669h	1	1	6	0	2	3 and 4
F9261h	0	1	3	0	2	6
M7524w	0	2	6	0	2	1
F3795w	1	1	6	0	2	1
M9231w	0	1	6	0	2	4
F5690h	0	1	6	0	2	1
M3609h	0	2	6	0	1	2
F8809w	0	2	6	0	2	1
F6914w	0	1	1	0	2	6
M3960h	0	2	3	0	2	1
M9287h	0	2	6	0	2	4
F2197w	0	1	3	0	2	4

Key: Previous rehabilitation: 0 = None; 1 = Physiotherapy; 2 = Chiropractic; 3 = Private massage, This episode; currently taking analgesia/NSAIDS?: 1 = Yes, 2 = No, This episode; treatment via GP: 1 = No contact; 2 = Advice only; 3 = oral or topical medication (Analg / NSAIDs) (+/- advice); 4 = oral medication + injection; 5 = injection only; 6 = referral only, This episode; number of GP injections, This episode; injection by non GP?: 1 = Yes, 2 = No, This episode; investigations: 1 = None; 2 = U/S; 3 = MRI; 4 = Xray; 5 = U/S and Xray, n/r = not recorded

13.3 C1. Patient reported measures: Pain

Table 13-5: Visual Analogue Scale (cm) (n = 76 at baseline)

Subject code	"On activity"	"At rest"	" At night over the last week"	Mean of 3 scores
F6847h	n/r	n/r	7.8	n/a
F7047h	n/r	n/r	8.1	n/a
M2146h	5.1	0.0	0.0	1.7
F5446h	2.5	0.0	4.4	2.3
F4417h	8.0	1.0	5.7	4.9
M0379h	4.9	2.5	1.2	2.9
F4583h	7.9	3.0	9.4	6.8
F7849h	1.1	1.1	2.6	1.6
M0825h	4.3	2.2	0.9	2.5
M0061h	8.2	6.6	8.0	7.6
F2520h	7.1	4.6	5.3	5.7
F8486h	2.4	1.5	1.6	1.8
F1005w	7.6	0.7	6.8	5.0
F6416h	7.4	5.1	5.9	6.1
F4764h	6.5	0.1	3.2	3.3
M4542h	8.4	2.4	8.4	6.4
M8878h	6.5	0.1	4.5	3.7
F0165h	6.6	4.7	7.3	6.2
F5367w	5.2	3.4	1.8	3.5
F7405h	1.7	1.0	0.2	1.0
M2747w	8.3	0.0	0.0	2.8
F1634w	6.5	4.9	5.4	5.6
F0738w	6.4	0.2	1.7	2.8
M0035w	6.7	2.3	3.5	4.2
F5503w	1.3	0.2	0.4	0.6
M9819h	7.0	0.0	0.8	2.6
M7535w	9.1	2.6	7.0	6.2
F0809h	6.3	4.1	5.3	5.2
F4113w	6.1	3.2	2.3	3.9
F9939h	7.7	1.1	8.1	5.6
M5703h	5.9	0.0	1.5	2.5
M1518h	7.4	2.8	7.4	5.9
F8774w	7.7	0.2	1.3	3.1
M8695h	0.2	1.2	0.1	0.5
F7304h	5.3	1.5	1.9	2.9
F2387h	3.1	0.8	0.8	1.6
F8396w	3.8	0.7	5.7	3.4
F4674w	7.2	0.7	3.7	3.9
M1243h	5.8	0.4	6.4	4.2
M2593w	2.8	0.3	4.4	2.5
M3288h	8.1	0.4	7.0	5.2
M0519h	6.6	1.3	7.4	5.1
M2608w	7.4	1.2	0.7	3.1
F8966w	7.9	5.8	8.2	7.3

F4943h	4.9	1.2	1.1	2.4
F5204w	3.3	0.6	0.1	1.3
M1662h	9.0	7.3	9.0	8.4
M5465w	2.8	0.7	0.7	1.4
M8305w	4.6	4.5	5.4	4.8
F0732w	2.0	0.4	2.0	1.5
F2122w	4.0	0.0	0.0	1.3
M2028W	8.1	2.9	8.4	6.5
F4142w	3.6	0.0	3.7	2.4
F9240h	6.4	1.5	9.1	5.7
F2573h	5.6	0.4	5.7	3.9
F4735h	2.1	0.3	4.7	2.4
F1965w	9.0	5.3	6.6	7.0
M6858w	6.0	1.1	3.6	3.6
F9830h	5.4	2.5	5.6	4.5
M2279h	8.6	8.7	8.9	8.7
M1543h	6.0	3.2	3.1	4.1
F1188h	6.5	4.6	6.8	6.0
M4091h	5.4	1.8	4.3	3.8
F5706h	4.7	1.8	2.8	3.1
F1669h	6.8	2.4	5.5	4.9
F9261h	8.5	2.3	5.6	5.5
M7524w	5.9	0.0	5.0	3.6
F3795w	7.6	3.3	7.6	6.2
M9231w	4.3	0.4	2.9	2.5
F5690h	5.5	0.2	4.0	3.2
M3609h	4.4	1.9	0.5	2.3
F8809w	8.1	0.9	4.6	4.5
F6914w	8.6	3.5	8.1	6.7
M3960h	7.2	1.2	7.8	5.4
M9287h	8.7	0.2	5.4	4.8
F2197w	3.6	7.1	1.3	4.0

Key: n/r = not recorded, n/a = not applicable

13.4 C2. Patient reported measures: Psychological symptoms

Table 13-6: 4DSQ: absolute score (n = 76 at baseline)

Subject code	Distress / absolute score	Depression / absolute score	Anxiety / absolute score	Somatisation / absolute score
F6847h	n/r	n/r	n/r	n/r
F7047h	14	1	4	12
M2146h	n/r	n/r	n/r	n/r
F5446h	8	1	2	2
F4417h	3	0	0	4
M0379h	0	0	0	1
F4583h	10	1	0	24
F7849h	21	5	2	12
M0825h	7	0	0	2
M0061h	8	1	1	8
F2520h	10	0	0	8
F8486h	5	0	0	5
F1005w	3	0	0	0
F6416h	5	0	2	7
F4764h	2	0	2	4
M4542h	11	0	0	3
M8878h	3	1	1	0
F0165h	3	0	0	1
F5367w	2	0	0	2
F7405h	2	0	0	6
M2747w	0	0	0	6
F1634w	32	10	2	16
F0738w	9	0	1	7
M0035w	22	2	3	10
F5503w	17	2	0	9
M9819h	5	0	0	1
M7535w	3	0	0	5
F0809h	30	12	11	13
F4113w	8	0	0	7
F9939h	0	0	0	0
M5703h	1	1	0	3
M1518h	19	5	4	12
F8774w	3	0	0	1
M8695h	5	0	0	5
F7304h	2	0	0	7
F2387h	3	0	0	11
F8396w	12	0	1	6
F4674w	0	0	0	4
M1243h	3	0	0	4
M2593w	3	0	0	3
M3288h	6	0	0	4
M0519h	20	3	10	9
M2608w	6	0	0	0

F8966w	7	0	0	6
F4943h	0	0	0	6
F5204w	5	0	0	4
M1662h	18	6	7	15
M5465w	0	0	0	1
M8305w	6	0	0	9
F0732w	2	0	0	1
F2122w	5	0	6	6
M2028W	2	0	0	3
F4142w	8	0	0	2
F9240h	29	7	15	13
F2573h	2	0	0	0
F4735h	8	0	5	13
F1965w	5	0	2	3
M6858w	20	4	6	8
F9830h	5	0	0	3
M2279h	8	0	1	8
M1543h	2	0	0	1
F1188h	7	0	0	7
M4091h	1	0	0	3
F5706h	14	4	3	11
F1669h	10	0	0	13
F9261h	13	2	4	5
M7524w	16	2	6	11
F3795w	18	1	1	4
M9231w	0	0	0	3
F5690h	6	0	0	11
M3609h	1	0	0	0
F8809w	9	0	0	5
F6914w	3	0	0	16
M3960h	7	0	2	4
M9287h	12	2	2	16
F2197w	5	0	0	2

Key: n/r = not recorded

Table 13-7: 4DSQ: categorised score (n = 76 at baseline)

Subject code	Distress / categorised	Depression / categorised	Anxiety / categorised	Somatisation / categorised
F6847h	n/r	n/r	n/r	n/r
F7047h	2	1	2	2
M2146h	n/r	n/r	n/r	n/r
F5446h	1	1	1	1
F4417h	1	1	1	1
M0379h	1	1	1	1
F4583h	1	1	1	3
F7849h	3	2	1	2
M0825h	1	1	1	1
M0061h	1	1	1	1
F2520h	1	1	1	1
F8486h	1	1	1	1
F1005w	1	1	1	1
F6416h	1	1	1	1
F4764h	1	1	1	1
M4542h	2	1	1	1
M8878h	1	1	1	1
F0165h	1	1	1	1
F5367w	1	1	1	1
F7405h	1	1	1	1
M2747w	1	1	1	1
F1634w	3	3	1	2
F0738w	1	1	1	1
M0035w	3	1	1	1
F5503w	2	1	1	1
M9819h	1	1	1	1
M7535w	1	1	1	1
F0809h	3	3	3	2
F4113w	1	1	1	1
F9939h	1	1	1	1
M5703h	1	1	1	1
M1518h	2	2	2	2
F8774w	1	1	1	1
M8695h	1	1	1	1
F7304h	1	1	1	1
F2387h	1	1	1	2
F8396w	2	1	1	1
F4674w	1	1	1	1
M1243h	1	1	1	1
M2593w	1	1	1	1
M3288h	1	1	1	1
M0519h	2	2	3	1
M2608w	1	1	1	1
F8966w	1	1	1	1
F4943h	1	1	1	1

F5204w	1	1	1	1
M1662h	2	3	2	2
M5465w	1	1	1	1
M8305w	1	1	1	1
F0732w	1	1	1	1
F2122w	1	1	2	1
M2028W	1	1	1	1
F4142w	1	1	1	1
F9240h	3	3	3	2
F2573h	1	1	1	1
F4735h	1	1	2	2
F1965w	1	1	1	1
M6858w	2	2	2	1
F9830h	1	1	1	1
M2279h	1	1	1	1
M1543h	1	1	1	1
F1188h	1	1	1	1
M4091h	1	1	1	1
F5706h	2	2	1	2
F1669h	1	1	1	2
F9261h	2	1	2	1
M7524w	2	1	2	2
F3795w	2	1	1	1
M9231w	1	1	1	1
F5690h	1	1	1	2
M3609h	1	1	1	1
F8809w	1	1	1	1
F6914w	1	1	1	2
M3960h	1	1	1	1
M9287h	2	1	1	2
F2197w	1	1	1	1

Key: n/r = not recorded

- Distress:
    - Score 0 to 10 = “not elevated”; coded as 1
    - Score 11 to 20 = “moderately elevated”; coded as 2
    - Score 21 to 32 = “strongly elevated”; coded as 3
- Depression:
    - Score 0 to 2 = “not elevated”; coded as 1
    - Score 3 to 5 = “moderately elevated”; coded as 2
    - Score 6 to 12 = “strongly elevated”; coded as 3
- Anxiety:
    - Score 0 to 3 = “not elevated”; coded as 1
    - Score 4 to 8 = “moderately elevated”; coded as 2
    - Score 9 to 24 = “strongly elevated”; coded as 3
- Somatisation:
    - Score 0 to 10 = “not elevated”; coded as 1
    - Score 11 to 20 = “moderately elevated”; coded as 2
    - Score 21 to 32 = “strongly elevated”; coded as 3

Table 13-8: 4DSQ: dichotomised score (n = 76 at baseline)

Subject code	Distress / categorised	Depression / categorised	Anxiety / categorised	Somatisation / categorised	Dichotomised score
F6847h	n/r	n/r	n/r	n/r	n/a
F7047h	2	1	2	2	1
M2146h	n/r	n/r	n/r	n/r	n/a
F5446h	1	1	1	1	0
F4417h	1	1	1	1	0
M0379h	1	1	1	1	0
F4583h	1	1	1	3	1
F7849h	3	2	1	2	1
M0825h	1	1	1	1	0
M0061h	1	1	1	1	0
F2520h	1	1	1	1	0
F8486h	1	1	1	1	0
F1005w	1	1	1	1	0
F6416h	1	1	1	1	0
F4764h	1	1	1	1	0
M4542h	2	1	1	1	1
M8878h	1	1	1	1	0
F0165h	1	1	1	1	0
F5367w	1	1	1	1	0
F7405h	1	1	1	1	0
M2747w	1	1	1	1	0
F1634w	3	3	1	2	1
F0738w	1	1	1	1	0
M0035w	3	1	1	1	1
F5503w	2	1	1	1	1
M9819h	1	1	1	1	0
M7535w	1	1	1	1	0
F0809h	3	3	3	2	1
F4113w	1	1	1	1	0
F9939h	1	1	1	1	0
M5703h	1	1	1	1	0
M1518h	2	2	2	2	1
F8774w	1	1	1	1	0
M8695h	1	1	1	1	0
F7304h	1	1	1	1	0
F2387h	1	1	1	2	1
F8396w	2	1	1	1	1
F4674w	1	1	1	1	0
M1243h	1	1	1	1	0
M2593w	1	1	1	1	0
M3288h	1	1	1	1	0
M0519h	2	2	3	1	1
M2608w	1	1	1	1	0
F8966w	1	1	1	1	0
F4943h	1	1	1	1	0

F5204w	1	1	1	1	0
M1662h	2	3	2	2	1
M5465w	1	1	1	1	0
M8305w	1	1	1	1	0
F0732w	1	1	1	1	0
F2122w	1	1	2	1	1
M2028W	1	1	1	1	0
F4142w	1	1	1	1	0
F9240h	3	3	3	2	1
F2573h	1	1	1	1	0
F4735h	1	1	2	2	1
F1965w	1	1	1	1	0
M6858w	2	2	2	1	1
F9830h	1	1	1	1	0
M2279h	1	1	1	1	0
M1543h	1	1	1	1	0
F1188h	1	1	1	1	0
M4091h	1	1	1	1	0
F5706h	2	2	1	2	1
F1669h	1	1	1	2	1
F9261h	2	1	2	1	1
M7524w	2	1	2	2	1
F3795w	2	1	1	1	1
M9231w	1	1	1	1	0
F5690h	1	1	1	2	1
M3609h	1	1	1	1	0
F8809w	1	1	1	1	0
F6914w	1	1	1	2	1
M3960h	1	1	1	1	0
M9287h	2	1	1	2	1
F2197w	1	1	1	1	0

Key: n/r = not recorded, n/a = not applicable

Dichotomised score: No categories with elevated = 0; 1 or more category with elevated = 1

13.5 C3. Patient reported measures: Function / Disability

Table 13-9: SPADI (%) (n = 76 at baseline)

Subject code	Total pain score	Total disability score	Total SPADI score
F6847h	77.5	14.3	37.3
F7047h	58.0	63.8	61.5
M2146h	54.0	28.0	41.0
F5446h	24.0	24.3	24.2
F4417h	76.0	75.0	75.4
M0379h	44.0	33.8	37.7
F4583h	72.0	82.0	77.0
F7849h	56.0	65.0	61.5
M0825h	42.0	11.3	23.1
M0061h	96.0	65.0	76.9
F2520h	84.0	52.5	64.6
F8486h	26.0	8.8	15.4
F1005w	58.0	52.5	54.6
F6416h	76.0	55.0	63.1
F4764h	32.0	32.5	32.3
M4542h	66.0	46.3	53.8
M8878h	68.0	40.0	50.8
F0165h	66.0	58.8	61.5
F5367w	56.0	27.5	38.5
F7405h	22.0	12.5	16.2
M2747w	36.0	5.0	16.9
F1634w	54.0	55.0	54.6
F0738w	40.0	43.8	42.3
M0035w	56.0	31.4	41.7
F5503w	14.0	8.8	10.8
M9819h	78.0	27.5	46.9
M7535w	70.0	66.3	67.7
F0809h	78.0	58.8	66.2
F4113w	38.0	8.8	20.0
F9939h	58.0	41.3	47.7
M5703h	46.0	28.8	35.4
M1518h	70.0	28.8	44.6
F8774w	64.0	53.8	57.7
M8695h	6.0	0.0	2.3
F7304h	30.0	37.5	34.6
F2387h	40.0	20.0	27.7
F8396w	54.0	23.1	43.9
F4674w	18.0	13.8	15.4
M1243h	50.0	41.3	44.6
M2593w	50.0	23.8	33.9
M3288h	62.0	31.3	43.8
M0519h	60.0	57.5	58.5
M2608w	46.0	23.8	32.3
F8966w	92.0	76.3	82.3

F4943h	42.0	20.0	28.5
F5204w	28.0	8.8	16.2
M1662h	96.0	86.3	90.0
M5465w	14.0	7.5	10.0
M8305w	54.0	23.8	35.4
F0732w	22.0	13.8	16.9
F2122w	8.0	0.0	3.1
M2028W	68.0	46.3	54.6
F4142w	26.0	16.3	20.0
F9240h	100.0	95.0	96.9
F2573h	58.0	30.0	40.8
F4735h	36.0	3.8	16.2
F1965w	88.0	63.8	73.1
M6858w	72.0	31.3	46.9
F9830h	22.0	17.5	19.2
M2279h	76.0	76.3	76.2
M1543h	60.0	46.3	51.5
F1188h	48.0	28.8	36.2
M4091h	46.0	2.5	19.2
F5706h	50.0	43.8	46.2
F1669h	68.0	55.0	60.0
F9261h	76.0	72.5	73.9
M7524w	72.0	38.8	51.5
F3795w	56.0	58.8	57.7
M9231w	46.0	23.8	32.3
F5690h	32.0	3.8	14.6
M3609h	56.0	33.8	42.3
F8809w	44.0	37.5	40.0
F6914w	74.0	66.3	69.2
M3960h	58.0	45.0	50.0
M9287h	68.0	40.0	50.8
F2197w	20.0	1.3	8.5

13.6 D1. Clinical measures: Strength

Table 13-10: Internal Rotation: raw and moment data (n = 76 at baseline)

Subject code	Distance elbow to radius (m)	Symptomatic side		Asymptomatic side	
		Mean Internal Rotation Force (N)	Mean Internal Rotation Moment (Nm)	Mean Internal Rotation Force (N)	Mean Internal Rotation Moment (Nm)
F6847h	0.24	29.3	7.0	41.8	10.0
F7047h	0.25	22.7	5.7	64.5	16.1
M2146h	0.29	38.1	11.0	52.8	15.3
F5446h	0.26	18.3	4.8	29.3	7.6
F4417h	0.27	19.1	5.2	35.2	9.5
M0379h	0.25	33.7	8.4	27.1	6.8
F4583h	0.24	8.8	2.1	24.9	6.0
F7849h	0.25	32.3	8.1	23.5	5.9
M0825h	0.29	44.0	12.8	44.0	12.8
M0061h	0.29	51.3	14.9	60.9	17.7
F2520h	0.24	14.7	3.5	79.2	19.0
F8486h	0.26	30.7	8.0	24.2	6.3
F1005w	0.25	44.0	11.0	72.6	18.2
F6416h	0.23	29.3	6.7	29.3	6.7
F4764h	0.28	19.8	5.5	29.3	8.2
M4542h	0.29	38.1	11.0	61.6	17.9
M8878h	0.28	44.0	12.3	55.0	15.4
F0165h	0.28	47.7	13.4	73.3	20.5
F5367w	0.24	50.6	12.1	80.7	19.4
F7405h	0.24	32.3	7.8	41.1	9.9
M2747w	0.32	164.7	52.7	181.3	58.0
F1634w	0.25	30.1	7.5	30.1	7.5
F0738w	0.26	37.4	9.7	33.7	8.8
M0035w	0.27	100.1	27.0	123.7	33.4
F5503w	0.28	38.1	10.7	63.8	17.9
M9819h	0.31	46.9	14.5	35.2	10.9
M7535w	0.28	103.7	29.0	100.2	28.1
F0809h	0.24	14.7	3.5	28.6	6.9
F4113w	0.25	30.1	7.5	33.7	8.4
F9939h	0.22	33.7	7.4	30.1	6.6
M5703h	0.27	75.5	20.4	65.3	17.6
M1518h	0.25	60.1	15.0	63.8	16.0
F8774w	0.22	29.3	6.4	28.6	6.3
M8695h	0.26	117.7	30.6	109.7	28.5
F7304h	0.25	68.9	17.2	96.1	24.0
F2387h	0.25	12.5	3.1	27.9	7.0
F8396w	0.24	44.0	10.6	72.6	17.4
F4674w	0.27	33.0	8.9	35.2	9.5
M1243h	0.28	100.9	28.3	120.7	33.8
M2593w	0.25	38.9	9.7	49.1	12.3

M3288h	0.26	125.7	32.7	149.0	38.7
M0519h	0.25	33.7	8.4	53.5	13.4
M2608w	0.28	70.4	19.7	106.0	29.7
F8966w	0.25	15.4	3.9	32.3	8.1
F4943h	0.25	33.7	8.4	46.9	11.7
F5204w	0.25	30.8	7.7	34.5	8.6
M1662h	0.27	19.1	5.2	60.1	16.2
M5465w	0.28	41.1	11.5	44.0	12.3
M8305w	0.29	60.1	17.4	55.0	16.0
F0732w	0.26	44.0	11.4	38.1	9.9
F2122w	0.22	24.2	5.3	29.3	6.4
M2028W	0.27	41.1	11.1	41.8	11.3
F4142w	0.26	35.2	9.2	34.5	9.0
F9240h	0.23	27.9	6.4	35.2	8.1
F2573h	0.21	44.0	9.2	35.9	7.5
F4735h	0.25	44.0	11.0	44.0	11.0
F1965w	0.25	15.4	3.9	44.0	11.0
M6858w	0.27	85.1	23.0	74.1	20.0
F9830h	0.24	54.3	13.0	50.6	12.1
M2279h	n/r	n/r	n/r	n/r	n/r
M1543h	0.28	44.0	12.3	50.6	14.2
F1188h	0.26	16.1	4.2	38.9	10.1
M4091h	0.31	37.4*	11.6*	44.0*	13.6*
F5706h	0.27	35.2	9.5	37.4	10.1
F1669h	0.26	39.6	10.3	38.9	10.1
F9261h	0.28	30.1	8.4	34.5	9.7
M7524w	0.28	76.3	21.4	70.4	19.7
F3795w	0.28	19.8	5.5	44.0	12.3
M9231w	0.27	48.4	13.1	79.2	21.4
F5690h	0.26	34.5	9.0	34.5	9.0
M3609h	0.27	30.1	8.1	56.5	15.3
F8809w	0.25	30.1	7.5	27.9	7.0
F6914w	0.23	20.5	4.7	44.0	10.1
M3960h	0.25	50.6	12.7	66.0	16.5
M9287h	0.28	38.1	10.7	76.3	21.4
F2197w	0.25	9.5	2.4	19.1	4.8

Key: m = metres, N = Newtons, Nm = Newton metres, n/r = not recorded, * = Symptoms on “Asymptomatic side” at time of testing

Table 13-11: External Rotation: raw and moment data (n = 76 at baseline)

Subject code	Distance elbow to radius /m	Symptomatic side		Asymptomatic side	
		Mean External Rotation Force / N	Mean External Rotation Moment / Nm	Mean External Rotation Force / N	Mean External Rotation Moment / Nm
F6847h	0.24	35.2	8.4	41.1	9.9
F7047h	0.25	9.3	2.3	52.1	13.0
M2146h	0.29	34.5	10.0	27.9	8.1
F5446h	0.26	26.4	6.9	24.9	6.5
F4417h	0.27	20.5	5.5	41.8	11.3
M0379h	0.25	32.3	8.1	35.9	9.0
F4583h	0.24	8.8	2.1	27.1	6.5
F7849h	0.25	32.3	8.1	32.3	8.1
M0825h	0.29	33.7	9.8	41.1	11.9
M0061h	0.29	46.9	13.6	45.5	13.2
F2520h	0.24	30.8	7.4	70.4	16.9
F8486h	0.26	30.1	7.8	30.1	7.8
F1005w	0.25	39.6	9.9	76.3	19.1
F6416h	0.23	30.8	7.1	24.2	5.6
F4764h	0.28	23.5	6.6	27.1	7.6
M4542h	0.29	30.8	8.9	46.9	13.6
M8878h	0.28	36.7	10.3	41.1	11.5
F0165h	0.28	35.2	9.9	58.7	16.4
F5367w	0.24	44.0	10.6	62.9	15.1
F7405h	0.24	30.8	7.4	36.7	8.8
M2747w	0.32	111.0	35.5	113.3	36.3
F1634w	0.25	16.9	4.2	23.5	5.9
F0738w	0.26	29.3	7.6	30.8	8.0
M0035w	0.27	92.4	24.9	111.0	30.0
F5503w	0.28	36.7	10.3	46.9	13.1
M9819h	0.31	41.8	13.0	47.7	14.8
M7535w	0.28	86.5	24.2	79.9	22.4
F0809h	0.24	8.8	2.1	25.7	6.2
F4113w	0.25	44.0	11.0	38.9	9.7
F9939h	0.22	19.1	4.2	19.8	4.4
M5703h	0.27	66.0	17.8	70.4	19.0
M1518h	0.25	51.3	12.8	55.7	13.9
F8774w	0.22	32.3	7.1	28.6	6.3
M8695h	0.26	73.3	19.1	74.8	19.4
F7304h	0.25	68.2	17.1	68.9	17.2
F2387h	0.25	11.7	2.9	12.5	3.1
F8396w	0.24	52.8	12.7	64.5	15.5
F4674w	0.27	31.5	8.5	41.1	11.1
M1243h	0.28	97.4	27.3	111.0	31.1
M2593w	0.25	50.6	12.7	62.3	15.6
M3288h	0.26	76.3	19.8	108.0	28.1
M0519h	0.25	35.2	8.8	63.1	15.8

M2608w	0.28	90.9	25.5	91.5	25.6
F8966w	0.25	11.7	2.9	30.1	7.5
F4943h	0.25	44.0	11.0	64.5	16.1
F5204w	0.25	35.2	8.8	35.2	8.8
M1662h	0.27	16.1	4.3	71.1	19.2
M5465w	0.28	44.0	12.3	78.5	22.0
M8305w	0.29	49.9	14.5	61.6	17.9
F0732w	0.26	41.1	10.7	35.2	9.2
F2122w	0.22	25.7	5.7	20.5	4.5
M2028W	0.27	69.7	18.8	55.0	14.9
F4142w	0.26	28.6	7.4	35.9	9.3
F9240h	0.23	14.7	3.4	35.2	8.1
F2573h	0.21	40.3	8.5	41.1	8.6
F4735h	0.25	41.1	10.3	63.1	15.8
F1965w	0.25	28.6	7.2	32.3	8.1
M6858w	0.27	111.0	30.0	127.7	34.5
F9830h	0.24	33.7	8.1	73.3	17.6
M2279h	n/r	n/r	n/r	n/r	n/r
M1543h	0.28	44.0	12.3	46.9	13.1
F1188h	0.26	16.1	4.2	33.7	8.8
M4091h	0.31	44.7*	13.9*	43.3*	13.4*
F5706h	0.27	33.7	9.1	53.5	14.4
F1669h	0.26	33.7	8.8	38.1	9.9
F9261h	0.28	8.8	2.5	35.2	9.9
M7524w	0.28	68.9	19.3	67.5	18.9
F3795w	0.28	24.9	7.0	44.0	12.3
M9231w	0.27	61.6	16.6	76.3	20.6
F5690h	0.26	34.5	9.0	40.3	10.5
M3609h	0.27	32.3	8.7	75.5	20.4
F8809w	0.25	11.7	2.9	15.4	3.9
F6914w	0.23	27.1	6.2	43.3	10.0
M3960h	0.25	36.7	9.2	68.2	17.1
M9287h	0.28	38.1	10.7	64.5	18.1
F2197w	0.25	8.8	2.2	8.8	2.2

Key: m = metres, N = Newtons, Nm = Newton metres, n/r = not recorded, * = Symptoms on “Asymptomatic side” at time of testing

Table 13-12: Scaption: raw and moment data (n = 76 at baseline)

Subject code	Distance GHJ to radius /m	Symptomatic side		Asymptomatic side	
		Mean Scaption Force / N	Mean Scaption Moment / Nm	Mean Scaption Force / N	Mean Scaption Moment / Nm
F6847h	0.52	24.2	12.6	33.7	17.5
F7047h	0.50	8.8	4.4	61.6	30.8
M2146h	0.60	8.8	5.3	26.4	15.8
F5446h	0.55	8.8	4.8	24.9	13.7
F4417h	0.53	0.0**	0.0**	30.8	16.3
M0379h	0.53	31.5	16.7	34.5	18.3
F4583h	0.52	5.9	3.1	19.1	9.9
F7849h	0.53	13.9	7.4	24.2	12.8
M0825h	0.54	22.0	11.9	35.2	19.0
M0061h	0.60	30.8	18.5	44.0	26.4
F2520h	0.50	14.7	7.4	62.3	31.2
F8486h	0.50	16.1	8.1	18.3	9.2
F1005w	0.48	8.8	4.2	44.0	21.1
F6416h	0.47	8.8	4.1	13.9	6.5
F4764h	0.59	8.8	5.2	11.0	6.5
M4542h	0.59	19.8	11.7	38.1	22.5
M8878h	0.59	25.7	15.2	30.8	18.2
F0165h	0.55	22.7	12.5	41.8	23.0
F5367w	0.51	35.2	18.0	47.7	24.3
F7405h	0.51	29.3	14.9	30.8	15.7
M2747w	0.57	106.7	60.8	100.3	57.2
F1634w	0.50	8.8	4.4	23.5	11.8
F0738w	0.54	12.5	6.8	22.0	11.9
M0035w	0.53	55.7	29.5	79.9	42.3
F5503w	0.51	13.2	6.7	44.0	22.4
M9819h	0.63	27.9	17.6	35.2	22.2
M7535w	0.51	44.0	22.4	44.0	22.4
F0809h	0.51	7.7	3.9	8.8	4.5
F4113w	0.53	27.9	14.8	32.3	17.1
F9939h	0.47	27.1	12.7	25.7	12.1
M5703h	0.51	46.9	23.9	46.9	23.9
M1518h	0.50	31.5	15.8	30.1	15.1
F8774w	0.49	8.8	4.3	18.3	9.0
M8695h	0.50	66.7	33.4	66.7	33.4
F7304h	0.51	31.5	16.1	66.0	33.7
F2387h	0.53	8.8	4.7	8.8	4.7
F8396w	0.50	13.9	7.0	37.4	18.7
F4674w	0.52	9.5	4.9	18.3	9.5
M1243h	0.52	61.6	32.0	69.7	36.2
M2593w	0.50	32.3	16.2	41.1	20.6
M3288h	0.50	77.0	38.5	68.2	34.1
M0519h	0.55	13.2	7.3	41.1	22.6
M2608w	0.55	52.1	28.7	85.8	47.2

F8966w	0.49	5.9	2.9	8.8	4.3
F4943h	0.51	28.1	14.3	44.0	22.4
F5204w	0.50	19.0	9.5	27.1	13.6
M1662h	0.55	6.6	3.6	18.3	10.1
M5465w	0.54	18.3	9.9	27.9	15.1
M8305w	0.54	32.3	17.4	31.5	17.0
F0732w	0.51	25.7	13.1	31.5	16.1
F2122w	0.49	17.6	8.6	11.7	5.7
M2028W	0.51	15.4	7.9	34.5	17.6
F4142w	0.48	9.5	4.6	24.9	12.0
F9240h	0.47	7.3	3.4	8.8	4.1
F2573h	0.47	8.8	4.1	8.8	4.1
F4735h	0.50	19.8	9.9	27.9	14.0
F1965w	0.48	8.8	4.2	27.1	13.0
M6858w	0.57	33.7	19.2	45.5	25.9
F9830h	0.49	16.1	7.9	35.2	17.2
M2279h	n/r	n/r	n/r	n/r	n/r
M1543h	0.56	18.3	10.2	32.3	18.1
F1188h	0.49	8.1	4.0	8.8	4.3
M4091h	0.58	19.1*	11.1*	22.7*	13.2*
F5706h	0.53	8.8	4.7	18.3	9.7
F1669h	0.50	14.7	7.4	22.7	11.4
F9261h	0.55	8.1	4.5	8.8	4.8
M7524w	0.55	42.5	23.4	71.1	39.1
F3795w	0.51	5.9	3.0	30.1	15.4
M9231w	0.52	24.9	12.9	44.0	22.9
F5690h	0.48	24.2	11.6	25.7	12.3
M3609h	0.53	28.6	15.2	50.6	26.8
F8809w	0.51	8.8	4.5	8.8	4.5
F6914w	0.47	8.8	4.1	8.8	4.1
M3960h	0.49	8.8	4.3	31.5	15.4
M9287h	0.52	33.0	17.2	41.1	21.4
F2197w	0.48	8.8	4.2	8.8	4.2

Key: m = metres, N = Newtons, Nm = Newton metres, n/r = not recorded, * = Symptoms on “Asymptomatic side” at time of testing, ** = Sub-threshold force

13.7 D2. Clinical measures: ROM

Table 13-13: Clinical measures: ROM (n = 76 at baseline)

Subject code	Active scaption ROM (°)		
	Symptomatic side		Asymptomatic side
	Limit of ROM	Point of symptom exacerbation by movement	Limit of ROM
F6847h	132.0	123.0	155.0
F7047h	150.0	60.0	162.0
M2146h	146.0	132.0	164.0
F5446h	163.0	135.0	179.0
F4417h	75.0	40.0	165.0
M0379h	145.0	135.0	155.0
F4583h	94.0	73.0	132.0
F7849h	168.0	155.0	163.0
M0825h	150.1	147.0	165.8
M0061h	146.0	105.0	146.0
F2520h	115.2	92.2	171.9
F8486h	155.0	138.0	159.0
F1005w	76.3	65.8	168.5
F6416h	137.0	70.0	157.0
F4764h	91.0	91.0	155.0
M4542h	155.0	55.0	155.0
M8878h	136.5	129.0	167.5
F0165h	148.0	100.0	157.0
F5367w	146.0	146.0	146.0
F7405h	154.0	120.0	154.0
M2747w	154.0	No symptoms during movement	154.0
F1634w	124.0	86.0	159.0
F0738w	166.0	150.0	166.0
M0035w	135.0	70.0	132.0
F5503w	124.0	104.0	155.2
M9819h	134.0	127.0	166.0
M7535w	147.0	107.0	164.0
F0809h	107.4	75.0	143.2
F4113w	152.0	130.0	152.0
F9939h	121.2	86.0	165.2
M5703h	146.0	132.0	161.0
M1518h	138.0	125.0	156.0
F8774w	99.5	96.0	153.2
M8695h	169.0	138.0	169.0
F7304h	131.0	121.0	168.4
F2387h	96.0	96.0	135.0
F8396w	111.5	106.0	163.7
F4674w	112.5	70.0	135.7
M1243h	108.0	108.0	161.0
M2593w	176.7	No symptoms during movement	176.7
M3288h	142.7	142.7	141.0 *
M0519h	150.0	70.0	167.0

M2608w	170.0	127.0	170.0
F8966w	70.0	65.0	155.0
F4943h	149.0	120.0	163.0
F5204w	157.0	132.0	157.0
M1662h	59.0	34.0	146.0
M5465w	162.0	91.0	155.0
M8305w	136.0	132.0	149.0
F0732w	160.0	105.0	169.0
F2122w	163.0	No symptoms during movement	163.0
M2028w	149.0	143.0	152.0
F4142w	115.0	94.0	157.0
F9240h	70.0	60.0	128.0
F2573h	138.0	128.0	138.0
F4735h	170.0	No symptoms during movement	170.0
F1965w	165.0	No symptoms during movement	168.0
M6858w	175.0	102.0	175.0
F9830h	105.0	92.0	151.0
M2279h	91.0	91.0	110.0
M1543h	93.0	93.0	153.0
F1188h	96.0	80.0	172.0
M4091h	149.0	119.0	153.0 *
F5706h	104.0	55.0	161.0
F1669h	164.0	No symptoms during movement	162.0
F9261h	74.0	74.0	162.0
M7524w	107.0	107.0	153.0
F3795w	84.0	84.0	158.0
M9231w	178.0	Pain remains same through range	174.0
F5690h	160.0	No symptoms during movement	160.0
M3609h	122.0	70.0	157.0
F8809w	101.5	101.5	164.0
F6914w	82.0	75.0	145.0
M3960h	77.0	Pain remains same through range	164.0
M9287h	174.0	118.0	174.0
F2197w	175.0	No symptoms during movement	174.0

Key: * = Symptoms on “Asymptomatic side” at time of testing

13.8 D3. Clinical measures: Scapular movement and control

Table 13-14: Scapular dyskinesis grading (n = 76 at baseline)

Subject code	Un-weighted trial grading	Weighted trial performed? 1=Y, 2=N	If weighted trial, what weight used? (kg)	Weighted trial grading
F6847h	2	1	0.5	2
F7047h	3	2	n/a	n/a
M2146h	3	1	1.0	3
F5446h	3	1	0.5	3
F4417h	3	1	0.5	3
M0379h	2	1	1.0	2
F4583h	1	2	n/a	n/a
F7849h	2	1	1.0	2
M0825h	2	1	0.5	2
M0061h	2	1	0.5	3
F2520h	3	2	n/a	n/a
F8486h	2	1	0.5	2
F1005w	1	2	n/a	n/a
F6416h	1	2	n/a	n/a
F4764h	1	2	n/a	n/a
M4542h	3	2	n/a	n/a
M8878h	1	1	0.5	2
F0165h	1	1	0.5	2
F5367w	2	1	0.5	2
F7405h	3	1	0.5	3
M2747w	2	1	1.0	2
F1634w	3	2	n/a	n/a
F0738w	2	1	0.5	2
M0035w	3	2	n/a	n/a
F5503w	2	1	0.5	2
M9819h	2	1	1.0	2
M7535w	1	2	n/a	n/a
F0809h	1	2	n/a	n/a
F4113w	1	1	0.5	2
F9939h	1	1	0.5	1
M5703h	3	1	0.5	3
M1518h	2	1	0.5	2
F8774w	2	1	0.5	2
M8695h	2	1	1.0	2
F7304h	2	2	n/a	n/a
F2387h	1	2	n/a	n/a
F8396w	3	2	n/a	n/a
F4674w	1	1	0.5	2
M1243h	2	1	1.0	2
M2593w	1	1	1.0	2
M3288h	2	1	0.5	2
M0519h	3	2	n/a	n/a

M2608w	2	1	0.5	2
F8966w	1	2	n/a	n/a
F4943h	1	1	0.5	1
F5204w	1	1	0.5	1
M1662h	3	2	n/a	n/a
M5465w	1	1	1.0	1
M8305w	2	1	0.5	2
F0732w	1	1	0.5	1
F2122w	1	1	0.5	1
M2028W	2	2	n/a	n/a
F4142w	2	1	0.5	2
F9240h	3	2	n/a	n/a
F2573h	2	1	0.5	2
F4735h	3	1	0.5	2
F1965w	3	1	1.0	3
M6858w	2	1	1.0	2
F9830h	2	2	n/a	n/a
M2279h	2	2	n/a	n/a
M1543h	2	2	n/a	n/a
F1188h	2	1	0.5	1
M4091h	1	1	1.0	1
F5706h	1	2	n/a	n/a
F1669h	2	1	0.5	2
F9261h	3	2	n/a	n/a
M7524w	3	2	n/a	n/a
F3795w	3	2	n/a	n/a
M9231w	2	1	1.0	3
F5690h	1	1	0.5	1
M3609h	1	2	n/a	n/a
F8809w	2	2	n/a	n/a
F6914w	1	2	n/a	n/a
M3960h	1	2	n/a	n/a
M9287h	1	2	n/a	n/a
F2197w	2	1	0.5	2

Key: n/a = not applicable

Scapular dyskinesis grading:

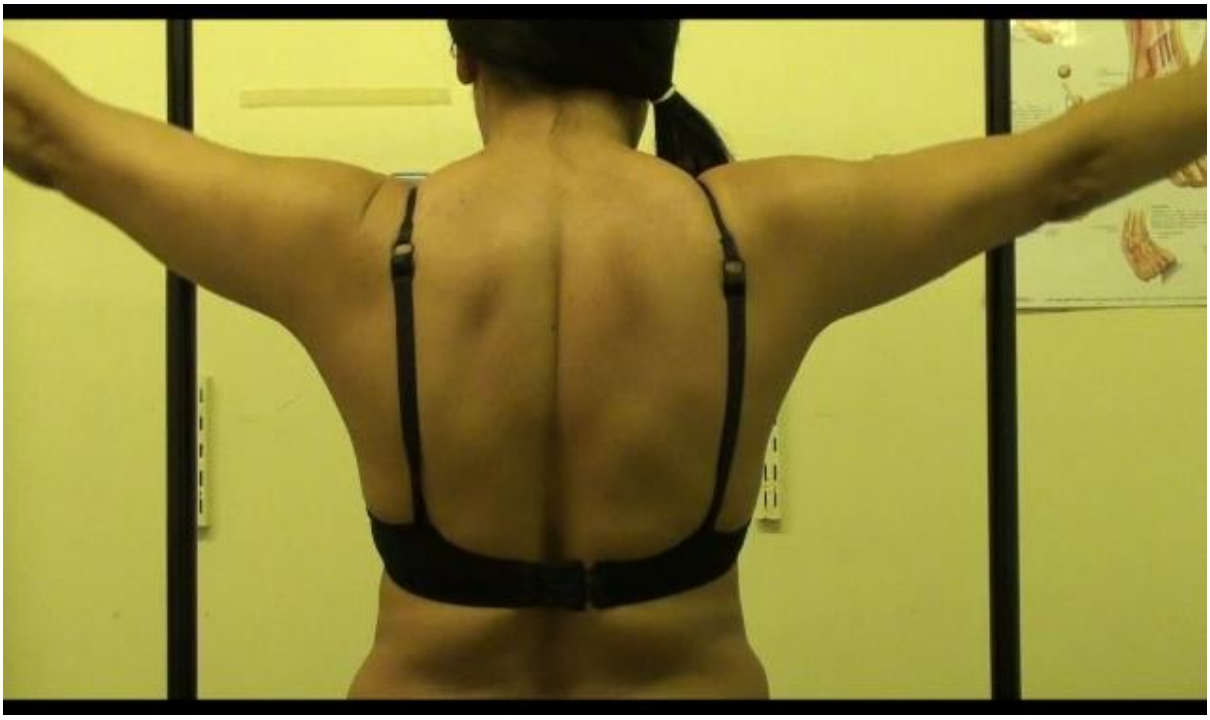
1 = “Normal”

2 = “Subtle abnormality”

3 = “Obvious abnormality”

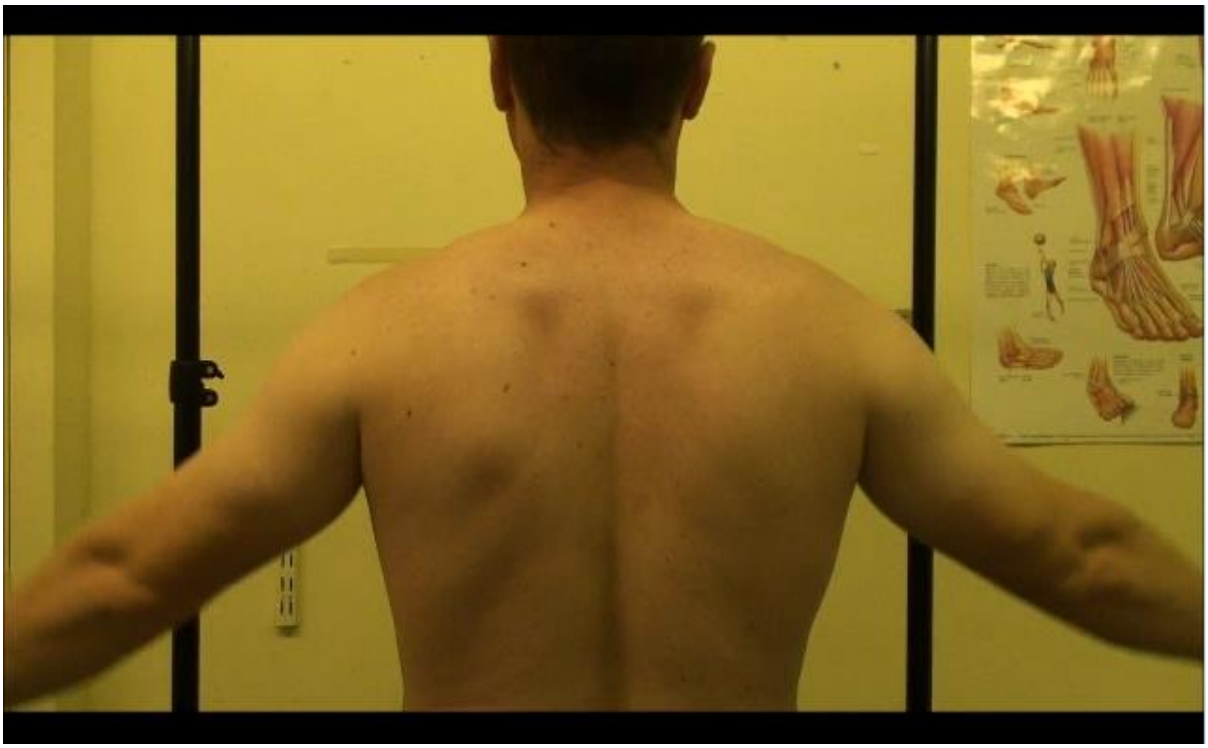
Examples of SDT grading

Figure 13-1: SDT trial for subject F9939



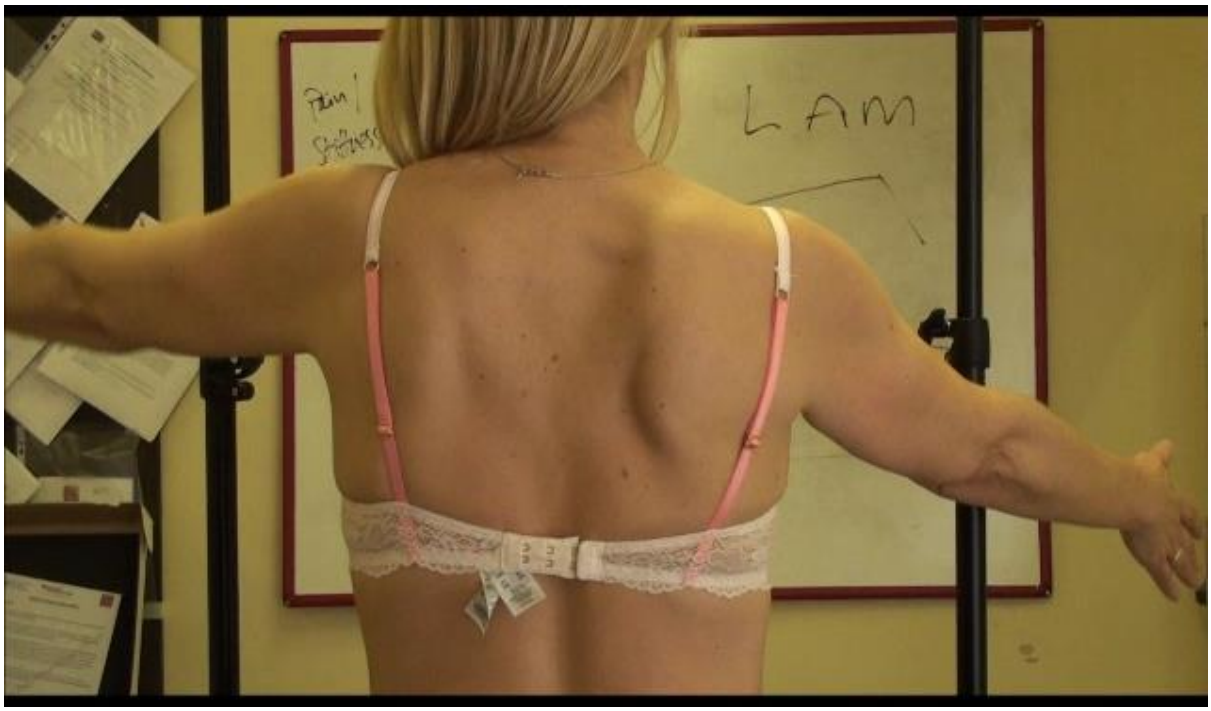
Subject F9939 demonstrating a scapular dyskinesis grading of 1 = “Normal”

Figure 13-2: SDT trial for subject M8695



Subject M8695 demonstrating a scapular dyskinesis grading of 2 = “Subtle abnormality” as evidenced by subtle prominence of the inferior angle of the scapular.

Figure 13-3: SDT trial for subject F3795



Subject F3795 demonstrating a scapular dyskinesis grading of 3 = “Obvious abnormality” as evidenced by obvious prominence of the inferior angle and medial border of the scapular.

13.9 E1. Structural pathology via imaging

Table 13-15: Sonographic diagnosis (findings for each component of the rotator cuff collapsed; composite grading) (n = 76 at baseline)

Subject code	Sonographic Diagnosis						
	Cuff pathology 1=Y, 2=N	Tendino pathic 1=Y, 2=N	Calcific deposit 1=Y, 2=N	Bursitis 1=Y, 2=N	PTT 1=Y, 2=N	FTT 1=Y, 2=N	Composite grading (Codeing below)
F6847h	2	2	2	1	2	2	1
F7047h	1	1	2	1	1	2	2
M2146h	1	1	2	1	2	2	2
F5446h	2	1	2	2	2	2	1
F4417h	1	1	2	1	2	1	2
M0379h	1	1	2	1	2	2	2
F4583h	1	1	2	2	2	2	1
F7849h	1	2	1	1	2	2	2
M0825h	2	2	2	1	2	2	1
M0061h	1	1	1	1	2	2	2
F2520h	2	2	2	1	2	2	1
F8486h	2	2	2	1	2	2	1
F1005w	1	1	2	1	2	2	2
F6416h	1	1	2	1	2	2	2
F4764h	2	2	2	2	2	2	0
M4542h	1	1	2	1	2	2	2
M8878h	2	2	2	2	2	2	0
F0165h	1	2	2	1	2	1	2
F5367w	1	1	2	2	2	2	1
F7405h	1	1	2	2	2	2	1
M2747w	2	2	2	1	2	2	1
F1634w	2	2	2	1	2	2	1
F0738w	1	1	2	1	2	2	2
M0035w	2	2	1	1	2	2	1
F5503w	1	1	2	2	2	2	1
M9819h	2	2	2	2	2	2	2
M7535w	1	1	2	2	2	2	1
F0809h	1	1	2	2	2	2	1
F4113w	1	1	2	2	2	2	1
F9939h	1	1	2	1	2	2	2
M5703h	2	2	2	2	2	2	0
M1518h	2	2	2	1	2	2	1
F8774w	2	2	2	1	2	2	1
M8695h	2	2	2	2	2	2	0
F7304h	1	2	1	2	2	2	1
F2387h	1	1	2	2	2	2	1
F8396w	2	2	2	2	2	2	0
F4674w	1	1	2	2	2	2	1
M1243h	1	1	2	1	2	2	2

M2593w	2	2	2	1	2	2	1
M3288h	1	1	2	1	2	2	2
M0519h	1	1	2	2	2	2	1
M2608w	2	2	2	2	2	2	0
F8966w	1	1	2	2	2	2	1
F4943h	2	2	2	1	2	2	1
F5204w	2	2	2	1	2	2	1
M1662h	1	1	2	1	2	2	2
M5465w	1	1	1	2	2	2	1
M8305w	2	2	2	2	2	2	0
F0732w	1	1	1	2	2	2	1
F2122w	2	2	2	2	2	2	0
M2028W	1	1	2	2	1	2	1
F4142w	1	1	2	1	2	2	2
F9240h	1	1	2	1	2	2	2
F2573h	2	2	2	1	2	2	1
F4735h	2	2	2	2	2	2	0
F1965w	1	1	2	2	2	2	1
M6858w	1	1	2	1	2	2	2
F9830h	2	2	2	2	2	2	0
M2279h	1	1	2	2	2	2	1
M1543h	1	1	2	1	2	2	2
F1188h	1	1	2	2	2	2	1
M4091h	1	1	2	2	2	2	1
F5706h	2	2	2	2	2	2	0
F1669h	2	2	2	1	2	2	1
F9261h	1	1	2	2	2	2	1
M7524w	1	1	2	1	2	2	2
F3795w	1	1	2	2	2	2	1
M9231w	1	1	2	n/r	2	1	n/a
F5690h	1	1	2	2	2	2	1
M3609h	2	2	2	1	2	2	1
F8809w	2	2	2	2	2	2	0
F6914w	1	1	2	1	2	2	2
M3960h	1	1	2	1	2	2	2
M9287h	1	1	2	1	2	2	2
F2197w	2	2	2	2	2	2	0

Key: n/r = not recorded, n/a = not applicable, PTT = partial thickness tear; FTT = full thickness tear
Composite grading: 0=no pathoplogy, 1 = 1 of bursal or cuff pathology, 2=both bursal and cuff pathology

Examples of using bilateral scanning as a mechanism to assist with tendinopathic differential diagnoses

Figure 13-4: Transverse view of Supraspinatus tendon on symptomatic side for subject M3609h

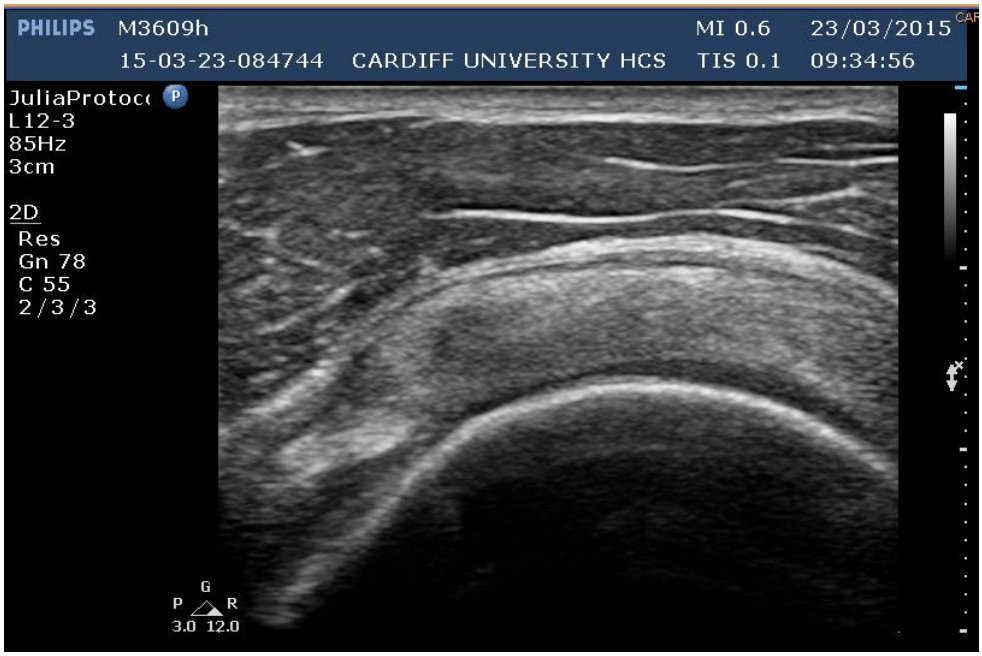
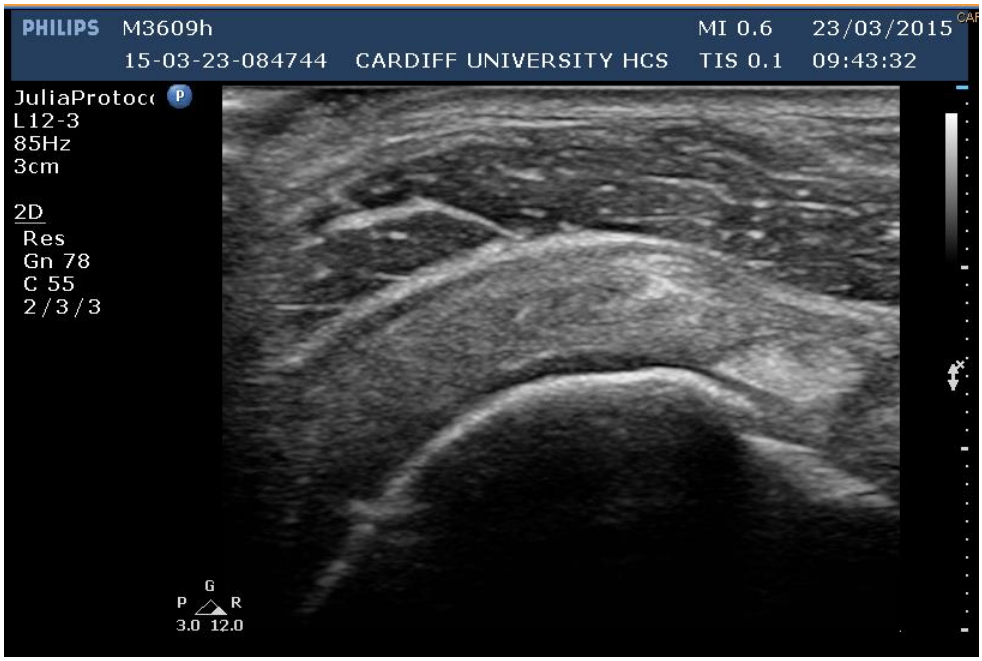


Figure 13-5: Transverse view of Supraspinatus tendon on asymptomatic side for subject M3609h



The transverse view of the Supraspinatus tendon on the symptomatic side for subject M3609h appears to be abnormally thickened, which in isolation can be considered characteristic of tendinopathy. However, the comparable image on the asymptomatic side demonstrates identical thickness. Therefore relative to the asymptomatic side there is no thickening of the Supraspinatus tendon on the symptomatic side.

Figure 13-6: Transverse view of Supraspinatus tendon on symptomatic side for subject F1965w

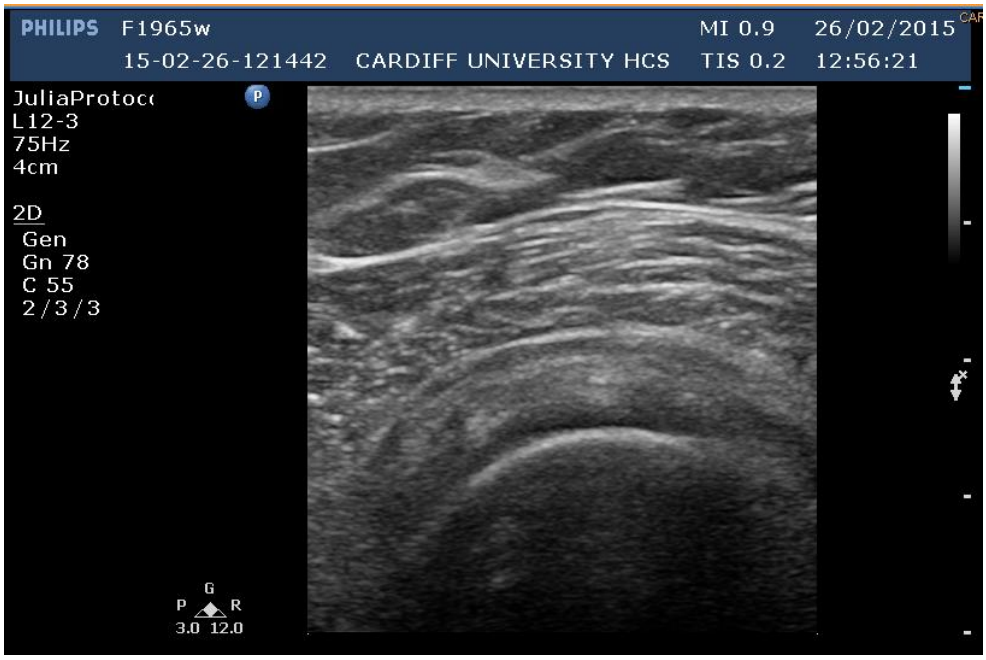
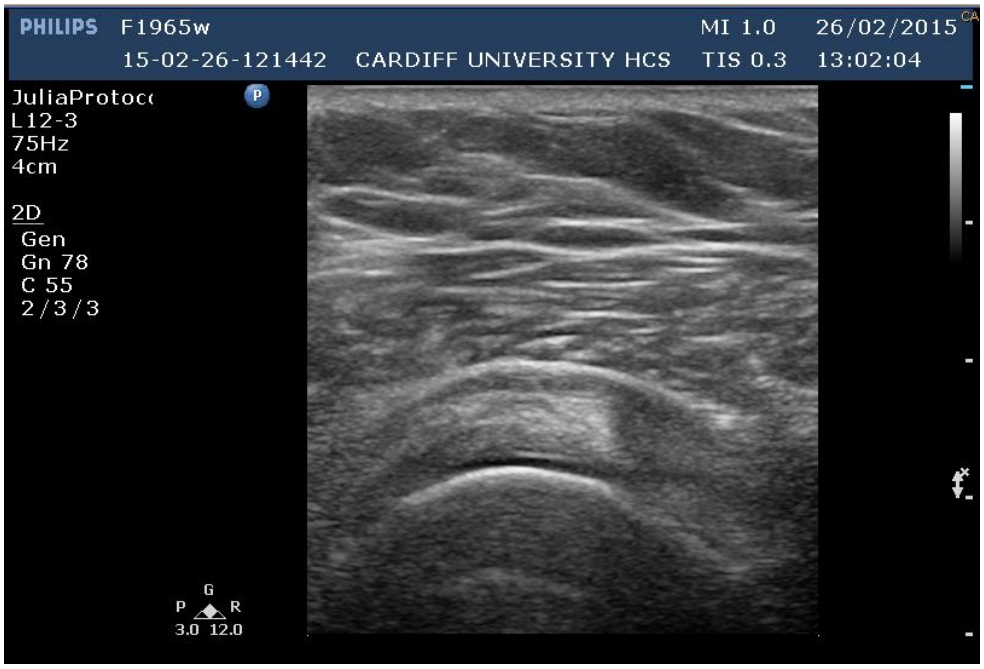


Figure 13-7: Transverse view of Supraspinatus tendon on asymptomatic side for subject F1965w



The images for subject F1965w demonstrate a high level of subcutaneous adipose tissue and this aligns with the high BMI of 39.9 Kg/m^2. The resulting attenuation of the ultrasound signal can impair image resolution, leading to reduced levels of differential diagnostic certainty. The use of bilateral scanning assisted with arriving at a diagnosis of tendinopathy as characterised by reduced echogenicity (increased darkness) on the symptomatic side, relative to the symptomatic side.

Figure 13-8: Transverse view of Supraspinatus tendon on symptomatic side for subject F9939h

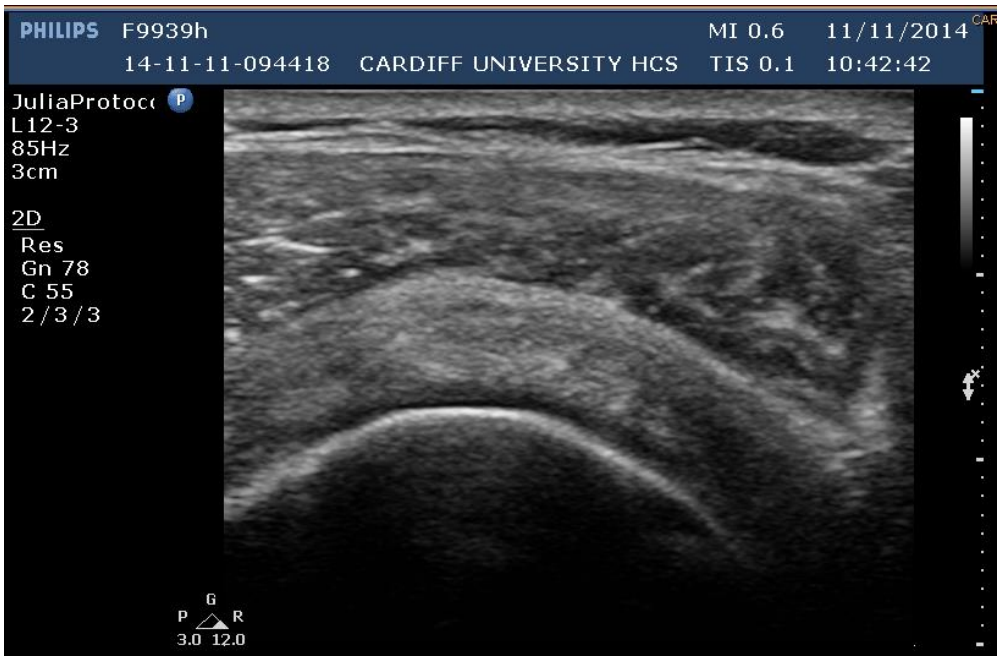
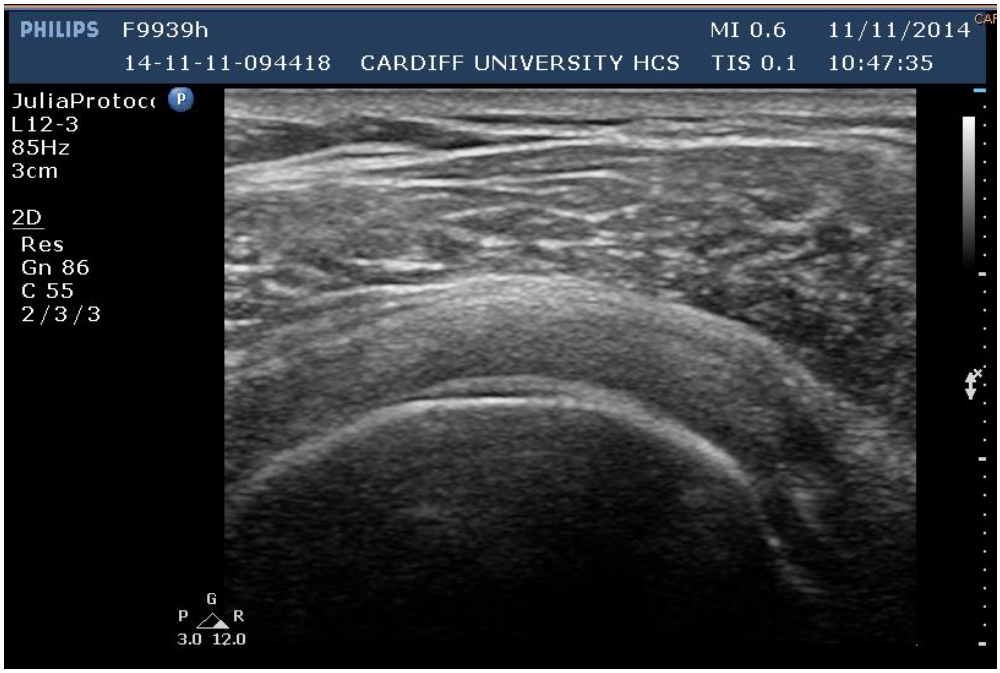


Figure 13-9: Transverse view of Supraspinatus tendon on asymptomatic side for subject F9939h



The transverse view of the Supraspinatus tendon on the symptomatic side for subject F9939h appears to be mildly heterogenous, although arguably within normal limits. However, the comparable image on the asymptomatic side demonstrates a high degree of homogeneity. Therefore relative to the asymptomatic side there is heterogeneity of the Supraspinatus tendon on the symptomatic side.

13.10 E2. Structural pathology via orthopaedic tests

Table 13-16: Orthopaedic tests (findings for each test; dichotomised for 3 or more +ve tests) (n = 76 at baseline)

Subject code	Hawkins-Kennedy 1 = +ve, 2=-ve	Neer & Walsh 1 = +ve, 2=-ve	Empty Can 1 = +ve, 2=-ve	Pain Active shoulder elevation 1 = +ve, 2=-ve	3 or more +ve tests
F6847h	1	1	n/r	1	1
F7047h	1	1	1	1	1
M2146h	2	2	1	1	0
F5446h	1	1	1	1	1
F4417h	1	1	1	1	1
M0379h	1	1	1	1	1
F4583h	1	1	1	1	1
F7849h	1	1	1	1	1
M0825h	1	1	1	1	1
M0061h	2	1	1	1	1
F2520h	1	1	1	1	1
F8486h	1	1	1	1	1
F1005w	1	1	1	1	1
F6416h	1	1	1	1	1
F4764h	1	1	1	1	1
M4542h	1	2	2	1	0
M8878h	2	2	1	1	0
F0165h	2	1	1	1	1
F5367w	2	1	2	1	0
F7405h	2	1	1	1	1
M2747w	2	2	2	2	0
F1634w	1	1	1	1	1
F0738w	2	2	1	1	0
M0035w	1	1	2	1	1
F5503w	2	2	1	1	0
M9819h	1	2	2	1	0
M7535w	1	1	1	1	1
F0809h	1	1	1	1	1
F4113w	1	2	2	1	0
F9939h	1	1	2	1	1
M5703h	2	1	1	1	1
M1518h	1	1	2	1	1
F8774w	2	1	1	1	1
M8695h	2	2	2	1	0
F7304h	2	1	1	1	1
F2387h	1	1	1	1	1
F8396w	1	2	2	1	0
F4674w	1	1	1	1	1
M1243h	2	1	1	1	1
M2593w	2	1	1	2	0

M3288h	1	1	2	1	1
M0519h	1	1	1	1	1
M2608w	1	2	1	1	1
F8966w	1	1	1	1	1
F4943h	2	1	1	1	1
F5204w	1	2	2	1	0
M1662h	1	1	1	1	1
M5465w	1	1	2	1	1
M8305w	1	1	1	1	1
F0732w	2	1	1	1	1
F2122w	2	2	2	2	0
M2028W	1	1	1	1	1
F4142w	1	1	1	1	1
F9240h	1	1	1	1	1
F2573h	2	1	1	1	1
F4735h	1	1	1	2	1
F1965w	1	1	2	2	0
M6858w	1	1	1	1	1
F9830h	2	1	1	1	1
M2279h	1	1	1	1	1
M1543h	1	1	1	1	1
F1188h	1	1	1	1	1
M4091h	1	2	2	1	0
F5706h	1	1	1	1	1
F1669h	1	1	1	2	1
F9261h	2	2	2	1	0
M7524w	1	1	2	1	1
F3795w	1	1	1	1	1
M9231w	1	1	1	1	1
F5690h	2	1	1	2	0
M3609h	2	1	1	1	1
F8809w	2	1	1	1	1
F6914w	1	1	1	1	1
M3960h	1	2	1	1	1
M9287h	2	1	2	1	0
F2197w	2	2	2	2	0

Key: n/r = not recorded

Hawkins-Kennedy, Neer & Walsh, Empty Can and Pain Active shoulder elevation: 1 = +ve, 2=-ve

3 or more +ve tests: 0 = no, 1 = yes

13.11 Number of weeks under treatment and discharge situation

Table 13-17: Number of weeks under treatment and discharge situation

Patient code	Total number of appointments	Number of weeks between 1st and last appointment	Discharge situation	Details
F6847h	1	0	B	Referred to Orthopaedics for injection
F7047h	5	9	A	
M2146h	2	3	A	
F5446h	8	35	A	
F4417h	11	21	A	
M0379h	5	11	A	
F4583h	7	26	A	
F7849h	4	10	A	
M0825h	5	12	A	
M0061h	8	14	A	
F2520h	5	15	A	
F8486h	5	12	A	
F1005w	5	19	A	
F6416h	2	3	C	
F4764h	6	16	A	
M4542h	9	32	A	
M8878h	4	9	A	
F0165h	5	13	A	
F5367w	4	7	A	
F7405h	4	8	A	
M2747w	4	20	A	
F1634w	8	42	B	Referred for psychological treatment
F0738w	1	0	C	
M0035w	2	13	C	
F5503w	4	12	A	
M9819h	2	1	C	
M7535w	8	19	A	
F0809h	4	5	A	
F4113w	6	15	A	
F9939h	5	19	C	
M5703h	4	19	C	
M1518h	7	17	A	
F8774w	2	2	A	
M8695h	2	9	A	
F7304h	4	13	A	
F2387h	3	7	A	
F8396w	10	28	A	
F4674w	7	27	A	
M1243h	5	9	A	
M2593	2	4	A	
M3288h	2	3	A	
M0519h	2	2	C	

M2608w	6	11	A	
F8966w	5	9	A	
F4943h	3	9	A	
F5204w	6	25	A	
M1662h	1	0	C	
M5465w	3	6	A	
M8305w	10	32	A	
F0732w	5	32	A	
F2122w	2	5	A	
M2028W	n/a	n/a	n/a*	
F4142w	7	17	A	
F9240h	8	26	A	
F2537h	4	9	A	
F4735h	5	11	A	
F1965w	1	0	C	
M6858w	1	0	C	
F9830h	2	8	A	
M2279h	5	15	A	
M1543h	3	8	B	Referred back to GP for injection
F1188h	1	0	C	
M4091h	5	17	A	
F5706h	8	20	A	
F1669h	3	7	A	
F9261h	6	22	A	
M7524w	3	7	A	
F3795w	13	37	A	
M9231w	8	34	B	Referred to Orthopaedics for surgery
F5690h	4	8	C	
M3609h	1	0	A	
F8809w	5	18	A	
F6914w	5	10	B	Referred to Orthopaedics for surgery
M3960h	4	6	A	
M9287h	2	2	A	
F2197w	3	32	C	

Key: A = Completed physiotherapy treatment and subsequently discharge from physiotherapy, B = Following physiotherapy treatment, care transferred to other speciality, C = Did not attend (DNA) or Unable to attend (UTA) and subsequently discharge from physiotherapy, n/a = not applicable, n/a* = Patient did not complete treatment for medical reasons unrelated to the shoulder

13.12 Outcome measure

Table 13-18: Outcome measure: Baseline to discharge to 3 months post-discharge

Subject code	Baseline OSS	Discharge OSS	Numerical change from Baseline to Discharge	Categorised change from Baseline to discharge	3 month follow up OSS	Numerical change from Baseline to 3 month follow up	Categorised change from Baseline to 3 month follow up	Categorised change from Baseline to Discharge to 3 month follow up
F6847h	33	29	-4	2	35	2	2	2a
F7047h	26	40	14	1	42	16	1	1a
M2146h	36	40	4	2	Missing	n/a	n/a	n/a
F5446h	37	43	6	1	48	11	1	1b
F4417h	27	44	17	1	44	17	1	1a
M0379h	37	40	3	2	42	5	1	3
F4583h	27	15	-12	3	14	-13	3	2b
F7849h	18	45	27	1	46	28	1	1a
M0825h	35	46	11	1	46	11	1	1a
M0061h	33	40	7	1	37	4	2	4
F2520h	31	37	6	1	43	12	1	1b
F8486h	43	41	-2	2	44	1	2	2a*
F1005w	19	44	25	1	29	10	1	1c
F6416h	22	36	14	1	36	14	1	1a
F4764h	36	45	9	1	43	7	1	1a
M4542h	25	45	20	1	46	21	1	1a
M8878h	33	48	15	1	48	15	1	1a
F0165h	32	31	-1	2	46	14	1	3
F5367w	33	45	12	1	46	13	1	1a
F7405h	39	44	5	1	41	2	2	4
M2747w	40	48	8	1	48	8	1	1a
F1634w	21	33	12	1	Missing	n/a	n/a	n/a

F0738w	40	44	4	2		46	6	1		3
M0035w	34	41	7	1		39	5	1		1a
F5503w	40	47	7	1		47	7	1		1a
M9819h	34	Missing	n/a	n/a		Missing	n/a	n/a		n/a
M7535w	26	47	21	1		42	16	1		1a
F0809h	26	35	9	1		39	13	1		1a
F4113w	35	41	6	1		41	6	1		1a
F9939h	27	37	10	1		Missing	n/a	n/a		n/a
M5703h	38	38	0	2		Missing	n/a	n/a		n/a
M1518h	34	42	8	1		46	12	1		1a
F8774w	33	48	15	1		48	15	1		1a
M8695h	41	46	5	1		44	3	2		4
F7304h	36	39	3	2		34	-2	2		2a
F2387h	36	36	0	2		Missing	n/a	n/a		n/a
F8396w	29	39	10	1		44	15	1		1b
F4674w	38	40	2	2		42	4	2		2a
M1243h	27	38	11	1		39	12	1		1a
M2593w	37	42	5	1		43	6	1		1a
M3288h	37	38	1	2		46	9	1		3
M0519h	29	46	17	1		46	17	1		1a
M2608w	36	45	9	1		47	11	1		1a
F8966w	16	41	25	1		38	22	1		1a
F4943h	36	45	9	1		48	12	1		1a
F5204w	40	40	0	2		45	5	1		3
M1662h	17	10	-7	3		19	2	2		2c
M5465w	41	45	4	2		48	7	1		3
M8305w	36	46	10	1		48	12	1		1a
F0732w	40	47	7	1		Missing	n/a	n/a		n/a
F2122w	42	46	4	2		43	1	2		2a
M2028W	37	Missing	n/a	n/a		Missing	n/a	n/a		n/a
F4142w	33	45	12	1		45	12	1		1a
F9240h	8	17	9	1		15	7	1		1a

F2573h	32	40	8	1		34	2	2	4
F4735h	41	43	2	2		45	4	2	2a
F1965w	22	45	23	1		47	25	1	1a
M6858w	34	40	6	1		Missing	n/a	n/a	n/a
F9830h	36	44	8	1		43	7	1	1a
M2279h	15	42	27	1		48	33	1	1b
M1543h	32	31	-1	2		37	5	1	3
F1188h	30	32	2	2		Missing	n/a	n/a	n/a
M4091h	40	42	2	2		45	5	1	3
F5706h	29	36	7	1		40	11	1	1a
F1669h	19	45	26	1		Missing	n/a	n/a	n/a
F9261h	17	46	29	1		48	31	1	1a
M7524w	25	32	7	1		26	1	2	4
F3795w	23	42	19	1		45	22	1	1a
M9231w	38	26	-12	3		31	-7	3	2b
F5690h	39	44	5	1		Missing	n/a	n/a	n/a
M3609h	34	39	5	1		Missing	n/a	n/a	n/a
F8809w	26	46	20	1		48	22	1	1a
F6914w	23	12	-11	3		16	-7	3	2b
M3960h	27	34	7	1		36	9	1	1a
M9287h	34	45	11	1		48	14	1	1a
F2197w	42	Missing	n/a	n/a		Missing	n/a	n/a	n/a

Key: n/a = not applicable

Change Baseline to discharge and Baseline to 3 month follow up:

1 = “Improved”

2 = “Same”

3 = “Worse”

Change to 3 month follow up:

1a = “Improvement” from baseline to discharge and “Improvement” maintained at 3 month post-discharge

1b = “Improvement” from baseline to discharge and further “Improvement” from discharge to 3 month post-discharge

1c = “Improvement” from baseline to discharge; “worsening” (according to MCIC) from discharge to 3 month post-discharge, but 3 month post-discharge level still “Improved” relative to baseline

2a = “Same” from baseline at any time point

2a* = “Same” from baseline at any time point; note baseline level meant that maximum OSS score required to demonstrate “Improvement”
2b = “Worse” from baseline to discharge and “Worse” compared to baseline maintained at 3 month post-discharge
2c = At discharge was “worse” compared to baseline; at 3 month post-discharge was within MCIC of baseline
3 = “Same” at discharge but at 3 month post-discharge had “Improved” from baseline
4 = “Improved” from baseline to discharge but within MCIC of baseline (i.e. same relative to baseline) at 3 month post-discharge

13.13 Testing for data normality for data reduction phase and differences between those lost to follow up and those where follow up data was available

13.13.1 A. Demographics: Age

Table 13-19: Kolmogorov-Smirnov and Shapiro-Wilk statistics for age variable

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	Degrees of freedom	Significance	Statistic	Degrees of freedom	Significance
Age	0.130	76	0.003	0.937	76	0.001

The Kolmogorov-Smirnov and Shapiro-Wilk statistics demonstrated that age was not normally distributed (Field 2009).

13.13.2 C1. Patient reported measures: Pain

Table 13-20: Kolmogorov-Smirnov and Shapiro-Wilk statistics for VAS variables

	Kolmogorov-Smirnov			Shapiro-Wilk		
VAS variable	Statistic	Degrees of freedom	Significance	Statistic	Degrees of freedom	Significance
On activity	0.099	74	0.071	0.951	74	0.006
At rest	0.167	74	<0.001	0.860	74	<0.001
At night	0.102	76	0.048	0.942	76	0.002
Mean	0.069	74	0.200	0.981	74	0.311

The Kolmogorov-Smirnov and Shapiro-Wilk statistics demonstrated that whilst pain "On activity" (assessed via Kolmogorov-Smirnov) and mean of 3 scores (assessed via both tests) were normally distributed, the other variables were not (Field 2009).

13.13.3 C3. Patient reported measures: Function / Disability

Table 13-21: Kolmogorov-Smirnov and Shapiro-Wilk statistics for SPADI variables

	Kolmogorov-Smirnov			Shapiro-Wilk		
SPADI variable	Statistic	Degrees of freedom	Significance	Statistic	Degrees of freedom	Significance
Total pain	0.079	76	0.200	0.985	76	0.505
Total disability	0.077	76	0.200	0.969	76	0.063
Total SPADI	0.082	76	0.200	0.983	76	0.409

The Kolmogorov-Smirnov and Shapiro-Wilk statistics demonstrated that all of the variables were normally distributed (Field 2009).

13.13.4 D1. Clinical measures: Strength

Table 13-22: Kolmogorov-Smirnov and Shapiro-Wilk statistics for strength variables

	Kolmogorov-Smirnov			Shapiro-Wilk		
Strength variable	Statistic	Degrees of freedom	Significance	Statistic	Degrees of freedom	Significance
IR mean moment	0.210	75	<0.001	0.773	75	<0.001
ER mean moment	0.184	75	<0.001	0.869	75	<0.001
Scap mean moment	0.172	75	<0.001	0.787	75	<0.001

The Kolmogorov-Smirnov and Shapiro-Wilk statistics demonstrated that none of the variables were normally distributed (Field 2009).

13.13.5 D2. Clinical measures: ROM

Table 13-23: Kolmogorov-Smirnov and Shapiro-Wilk statistics for ROM variables

	Kolmogorov-Smirnov			Shapiro-Wilk		
ROM variable	Statistic	Degrees of freedom	Significance	Statistic	Degrees of freedom	Significance
Symptomatic side: limit of ROM	0.132	66	0.006	0.950	66	0.010
Symptomatic side: point of symptom exacerbation by movement	0.098	66	0.185	0.971	66	0.128

The Kolmogorov-Smirnov and Shapiro-Wilk statistics demonstrated that whilst point of symptom exacerbation by movement was normally distributed, limit of ROM was not (Field 2009).

13.13.6 Outcome measure: OSS at baseline

Table 13-24: Kolmogorov-Smirnov and Shapiro-Wilk statistics for OSS at baseline

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	Degrees of freedom	Significance	Statistic	Degrees of freedom	Significance
OSS at baseline	0.152	76	<0.001	0.936	76	0.001

The Kolmogorov-Smirnov and Shapiro-Wilk statistics demonstrated that OSS at baseline was not normally distributed (Field 2009).

13.14 Variable values for patients lost to follow

Table 13-25: Lost to follow up at discharge

Patient	Baseline OSS	Age	Symptom duration	4DSQ	Total SPADI score	Sx ER mean moment	Unloaded SDT	U/S pathology
M2028w	37	58	1	0	54.6	18.8	2	1
F2197w	42	27	4	0	8.5	2.2	2	0
M9819h	34	29	2	0	46.9	13.0	2	0

Key: Symptom duration: 1 = 0-3 months; 2 = 3-6 months; 3 = 6-12 months; 4 = 12-24 months; 5 = More than 24 months, 4DSQ: No categories with elevated = 0; 1 or more category with elevated = 1 , Unloaded SDT: 1 = "Normal", 2 = "Subtle abnormality", 3 = "Obvious abnormality", U/S pathology: 0=no pathoplogy, 1 = 1 of bursal or cuff pathology, 2=both bursal and cuff pathology

Table 13-26: Lost to follow up at 3 months post discharge

Patient	Baseline OSS	Age	Symptom duration	4DSQ	Total SPADI score	Sx ER mean moment	Unloaded SDT	U/S pathology
M2146h	36	54	5	n/a	41.0	10.0	3	2
F1634w	21	30	3	1	54.6	4.2	3	1
F9939h	27	50	3	0	47.7	4.2	1	2
M5703h	38	49	3	0	35.4	17.8	3	0
F2387h	36	66	2	1	27.7	2.9	1	2
F0732w	40	49	3	0	16.9	10.7	1	1
M6858w	34	34	4	1	46.9	30.0	2	2
F1188h	30	38	3	0	36.2	4.2	2	1
F1669h	19	18	5	1	60.0	8.8	2	1
F5690h	39	19	3	1	14.6	9.0	1	1
M3609h	34	40	5	0	42.3	8.7	1	0

Key: Symptom duration: 1 = 0-3 months; 2 = 3-6 months; 3 = 6-12 months; 4 = 12-24 months; 5 = More than 24 months, 4DSQ: No categories with elevated = 0; 1 or more category with elevated = 1, Unloaded SDT: 1 = "Normal", 2 = "Subtle abnormality", 3 = "Obvious abnormality", U/S pathology: 0=no pathoplogy, 1 = 1 of bursal or cuff pathology, 2=both bursal and cuff pathology